HUMAN RELATIONS

A Quarterly Journal of Studies towards the Integration of the Social Sciences

THE TAVISTOCK INSTITUTE OF HUMAN RELATIONS LONDON, ENGLAND.

THE RESEARCH CENTER FOR GROUP DYNAMICS ANN ARBOR, MICH., U.S.A.

Volume VII. Number 4. 1954 Pages 401—520.

EDITORIAL COMMITTEE

TAVISTOCK INSTITUTE OF HUMAN RELATIONS

ELLIOTT JAQUES
B. S. MORRIS
ERIC TRIST
A. T. M. WILSON

RESEARCH CENTER FOR GROUP DYNAMICS

DORWIN CARTWRIGHT
JOHN R. P. FRENCH, Jnr.
JAY M. JACKSON
RONALD LIPPITT
IAN C. ROSS
ALVIN ZANDER

Managing Editor
JOHN HARVARD-WATTS
Tavistock Publications Ltd.
2 Beaumont Street, London, W.1.

ADVISORY EDITORIAL BOARD

BRITAIN

H. V. DICKS
Tavistock Clinic, London

MEYER FORTES
Department of Anthropology, University of Cambridge

RONALD HARGREAVES
Mental Health Section
World Health Organisation

C. A. MACE
Department of Psychology, Birkbeck College, University of London

ALEC RODGER
Department of Psychology, Birkbeck College, University of London

T. FERGUSON RODGER
Department of Psychiatry, University of Glasgow

E. A. SHILS
Department of Sociology, University of Chicago

T. S. SIMEY
Department of Social Science, University of Liverpool

J. D. SUTHERLAND
Tavistock Clinic, London

D. W. WINNICOTT
Paddington Green Children's Hospital, London

NORTH AMERICA

GABRIEL A. ALMOND
Center of International Studies, Princeton University.

ROBERT F. BALES
Department of Social Relations, Harvard University.

STUART W. COOK
Department of Psychology, New York University.

LEONARD S. COTTRELL, Jr.
The Russell Sage Foundation, New York.

RICHARD S. CRUTCHFIELD
Institute of Personality Assessment and Research, University of California.

LEON FESTINGER
Department of Psychology, University of Minnesota.

JACOB E. FINESINGER
School of Medicine, University of Maryland.

CARL I. HOVLAND
Department of Psychology, Yale University.

GEORGE KATONA
Institute for Social Research, University of Michigan.

DANIEL KATZ
Department of Psychology, University of Michigan.

CHARLES P. LOOMIS
Department of Sociology and Anthropology, Michigan State College.

MARGARET MEAD
Institute for Intercultural Studies, New York.

GEORGE SASLOW
Department of Neuropsychiatry, Washington University School of Medicine.

HERBERT THELEN
School of Education, University of Chicago.

WILLIAM F. WHYTE
School of Industrial and Labor Relations, Cornell University.

HUMAN RELATIONS

Studies towards the Integration of the Social Sciences

Vol. VII, No. 4 Pages 401–520

Page

STANLEY SCHACHTER, JOSEF NUTTIN, CECILY DE MONCHAUX, PAUL H. MAUCORPS, DIEDRICH OSMER, HUBERTUS DUIJKER, RAGNAR ROMMETVEIT, and JOACHIM ISRAEL. Cross-Cultural Experiments on Threat and Rejection 403

F. KRÄUPL TAYLOR. The Three-Dimensional Basis of Emotional Interactions in Small Groups. I 441

ORVILLE G. BRIM, Jr. The Acceptance of New Behavior in Child-Rearing 473

RAYMOND B. CATTELL and GLEN F. STICE. Four Formulae for Selecting Leaders on the Basis of Personality 493

LEONARD BERKOWITZ. Group Standards, Cohesiveness, and Productivity 509

THE TAVISTOCK INSTITUTE OF HUMAN RELATIONS
LONDON

THE RESEARCH CENTER FOR GROUP DYNAMICS
ANN ARBOR, MICH.

Published by
TAVISTOCK PUBLICATIONS LTD.
2 BEAUMONT STREET, LONDON, W.1.

1954

HUMAN RELATIONS

SUBSCRIPTIONS

In the United Kingdom, the British Commonwealth of Nations (except Canada), Europe, Middle East, and Far East: 30/– per volume (4 issues) and 8/6 per issue. Postage free.

Subscriptions should be addressed to the Managing Editor, TAVISTOCK PUBLICATIONS LTD., 2 Beaumont Street, London, W.1.

Cheques should be made payable to Tavistock Publications Ltd.

* * *

In the United States, Canada, and Latin America: $6.00 per volume (4 issues) and $1.75 per issue. Postage free. Special Rates for extended subscriptions.

Two years, $10.00

Three years, $14.00

Subscriptions should be addressed to HUMAN RELATIONS, The Research Center for Group Dynamics, University of Michigan, Ann Arbor, Mich., U.S.A.

* * *

Back Numbers. Copies may be obtained from the above addresses at the rates quoted.

EDITORIAL NOTICE

MSS. of papers offered for publication should be submitted as follows:

In the United Kingdom, the British Commonwealth of Nations (except Canada), Europe, Middle East, and Far East, to the Managing Editor, *Human Relations*, Tavistock Publications Ltd., 2 Beaumont Street, London, W.1.

In the United States, Canada, and Latin America, to the Secretary, Editorial Committee, *Human Relations*, The Research Center for Group Dynamics, University of Michigan, Ann Arbor, Mich., U.S.A.

Authors of papers published in *Human Relations* will receive, free of charge, one copy of the issue concerned, and 50 offprints of their paper. Further copies of the offprints can be obtained by order in advance of publication from the above sources.

CROSS-CULTURAL EXPERIMENTS ON THREAT AND REJECTION[1]

A Study of the Organization for Comparative Social Research

STANLEY SCHACHTER, JOSEF NUTTIN, CECILY DE MONCHAUX, PAUL H. MAUCORPS, DIEDRICH OSMER, HUBERTUS DUIJKER, RAGNAR ROMMETVEIT, AND JOACHIM ISRAEL

The present paper reports the first study of the Organization for Comparative Social Research. This organization is made up of social scientists representing the following research and educational institutions in western Europe: the University of Louvain, Belgium; the London School of Economics, and University College, London, England; the Centre d'Études Sociologiques, and the National Institute of Statistics and Economic Studies, France; the University of Frankfurt, and the Institut für Demoscopie, Germany; the University of Amsterdam, and the University of Nijmegen, Holland; the Institute for Social Research, Norway; and the University of Stockholm, Sweden.

The purposes of the OCSR have been 1. to encourage cooperation among the social scientists of different countries; 2. to increase the training facilities for sociologists and psychologists in Europe in the newer techniques of research on human problems, and 3. to explore the potentialities and difficulties of cross-cultural research on problems within the fields of social psychology and sociology.

The Institute for Social Research of Oslo initiated this program of cross-cultural research and training with the cooperation of visiting Fulbright professors. This Institute supported the first planning stages of the project,

1. The group experiment which this paper reports was conducted under the auspices of the Organization for Comparative Social Research. Working with Dr. Stanley Schachter, as Research Coordinator, were two research groups in each of seven countries, composed as follows :—BELGIUM : University of Louvain : Josef Nuttin (Research Director), Rafael Kluppels, Valère Lamsens ; FRANCE : Centre d'Études Sociologiques : Paul Maucorps (Research Director), Jacques van Bockstaele ; GERMANY : Institut für Sozialforschung : Diedrich Osmer (Research Director), Volker van Hagen, Werner Mangold, Christoph Oehler, Inge Rauter, Peter Schönbach ; NETHERLANDS : University of Amsterdam : Hubertus Duijker (Research Director), Mauritz Mulder ; NORWAY : Institute for Social Research : Ragnar Rommetveit (Research Director), Gerdt Henrik Vedeler ; SWEDEN : University of Stockholm : Joachim Israel (Research Director), Irene Steg ; UNITED KINGDOM : University College, London : Cecily de Monchaux (Research Director), Sylvia Shimmin, Kenneth Miller, Godfrey Gardner, Brian Allt.

which included two international seminars in Oslo and pilot studies in the cooperating countries. This preliminary work was followed by a third conference at the University of Amsterdam, where the participants organized themselves into a permanent body and agreed upon the first phases of their joint research plans. During the early part of this period of organization and formalization (1950–51), Dr. Herbert Hyman served as consultant and coordinator to the group. He was followed in this role by Dr. Daniel Katz for the period 1951–52.

At these meetings the group agreed upon a program of replicated researches on the relationships of threat and conformity, and, more particularly, on attitudes toward deviant group members under conditions of threat. It was agreed that the problem should be studied by means of an attitude survey and a group experiment. In each participating country two research groups were appointed, one to conduct the survey and one to conduct the experiment. The present paper reports the results of the group experiment. Results of the survey will be reported in later publications.

Active research was conducted during the period 1952–53 with the support of a grant from the Ford Foundation. During this time and the following period of analysis and write-up, Dr. Stanley Schachter acted as consultant and research coordinator for the group experiment team.

CROSS-CULTURAL RESEARCH

In most areas of present day social science it is impossible to distinguish universal from culturally determined relationships. Most studies are concerned with particular subgroups within a culture. Replications are rare, and, with a few exceptions, there have been almost no attempts to replicate studies on different subgroups within a culture or in different cultures. Since cross-cultural replication is a difficult process, sharply limited by the availability of funds and suitably trained personnel, it is scarcely surprising that there have been few such studies. The situation is understandable but unfortunate, for, until a large body of such studies is available, it will remain difficult to distinguish the limits of generalizability of specific propositions or narrowly determined lawful relationships.

The present paper presents the first exploratory steps of a program of replicated research in western Europe. It is exploratory in the sense of exposing and attempting tentative solutions to the methodological and interpretive problems unique to cross-cultural replications.

The unique and central problem of replicated research is clearly indicated in the question, "If there are differences in the results of the several experiments, how are they to be explained?" Such differences may be attributable to many factors. They may be due to experimental artifacts, such as failure to produce the independent variables, or poor translations of measuring instruments; they may indicate cultural differences; or they may indicate

that the original hypotheses were inadequate, in which case an alternative theory may be able to reconcile apparently divergent results. For many differences between experiments any of these may be a plausible explanation. If a particular set of replications is to be more than a vehicle for bickering, it must be possible, to some degree, to determine which of these factors is responsible for a difference. The methodology unique to replicated research must, then, be a protective methodology designed to provide the evaluative data which will allow choice among the possible interpretations of potential differences.

More than these protective-evaluative measures, however, is desirable. Though it may be personally enlightening to learn that the difference between two experiments is attributable to an experimental artifact, such a finding contributes little to knowledge. It is desirable, of course, to eliminate such factors and to standardize to the maximum the several experiments.

The success of the attempt to standardize is governed to a high degree by training and meticulousness. However, these qualities alone are not always sufficient. Though it may be possible to create experiments which, in execution, are virtual images of one another, this alone is no guarantee that the several experiments produce conceptually equivalent situations. An effective manipulation or experimental setting in one nation may be completely unsuccessful in another. For example, an experimental setting centering around a discussion of skiing would probably arouse immediate interest among subjects in Norway but would be rather remote and tangential to the interest of most subjects in Holland. Though the experimental settings might be identical in the two countries, the two experiments would differ considerably.

Because the problems of this sort of research are still novel, in this paper we shall devote considerable attention to the methodological and interpretive details of the over-all study. For some of the problems mentioned briefly above we offer tentative solutions; for others, documentation.

THE PROBLEM

The problem on which the experiments of the several participating countries focused was concerned with the relationship of threat to tolerance and conformity. Reactions to a deviate from a group norm are studied in the experimentally produced setting of a threatened group goal. Threat is defined as the possibility that a goal will not be achieved, and it is assumed that the magnitude of threat is a function of the valence or desirability of the goal and the probability that the goal will not be reached. Thus, when a goal is highly valent and conditions are such that it is improbable that the goal will be reached, threat is high; when valence is low and it is likely that the goal will be reached, threat is low.

In the course of a variety of studies, Festinger, Back, Schachter, and others,

(1, 2, 3, 4, 9) have developed a theory of social influence and communication pertinent to the present study. Briefly, they postulate that if differences of opinion exist within a group, pressures to uniformity will arise 1. if the validity or "reality" of an opinion depends upon agreement with the group, or 2. if locomotion toward a group goal will be facilitated by uniformity within the group.

When differences of opinion exist, pressures to uniformity will be manifested in the development of three tendencies, (a) a tendency to change one's own opinion to conform to that of other group members, (b) a tendency to change the opinion of others, and (c) a tendency to redefine the boundaries of the group so as to exclude those holding deviate opinions. It has been demonstrated that the magnitude of these several tendencies will vary with 1. the cohesiveness of the group (1, 9), 2. the relevance of the issue (9), and 3. the degree and pattern of discrepancy which exists within the group (4, 9).

It can be seen that our problem concerning the relationship of threat to tolerance is closely related to this theory. Threat has been defined in terms of group locomotion. It would seem plausible to suggest that pressures to uniformity on issues relevant to group locomotion should vary with the desirability of the group goal and the probability of goal achievement. The present study examines the relationships between goal valence, probability of goal achievement, and the tendencies to change others and to reject deviates.

METHODOLOGY

This section is divided into two parts: 1. a description of the specific experiment run commonly by each of the participating countries; 2. a description of the procedures adopted to insure some degree of standardization among the various experiments and of the measures employed to check on the comparability of the several experiments.

THE COURSE OF THE EXPERIMENT

The experiment was conducted in the guise of the first meeting of a boys' aviation club. The subjects were all school boys who had previously been recruited from their classrooms. The recruiter had represented himself as coming from a fictitious Institute of Aviation which was interested in forming clubs for boys interested in aviation. These clubs were glamorously described in terms of the opportunities they would offer to meet pilots, build models, see aviation movies, visit airfields, fly, etc. In all countries the response of the boys to the recruiter was enthusiastic, and in the several countries an average of 84 per cent of the boys addressed asked to become members of these clubs (Range 71 per cent–99 per cent).

A week or two after volunteering, the boys were requested by letter to come to the first meeting of their club. Each club was made up of six or seven boys previously unknown to one another. In each country the meeting was conducted in a room decorated with airline posters, models and an assortment of massive and exotic bits of aeronautical junk. After a brief introduction, the forty-five minutes of the meeting were devoted to a "glamour" period designed to intrigue the boys and put them at their ease. During this time they played with working models of jet and benzine motors, threw gliders to one another, and so on.

Following this interval, the experimenter called the boys together, had them choose a name for their club, and repeated with additional flourishes the promises of the recruiter's speech. He continued with the manipulation of valence and probability. In the several conditions, goals of varying desirability were established, and the probability that a particular club would reach its goal was varied so that some groups had a very good chance and others a poor chance.

To decide among the various clubs, a competition was announced in which the clubs building the best model airplanes would achieve the goal. The experimenter described in detail five different models for which kits were available, and told the boys that they would work together and could choose one of these models to build. Four of the models were extremely attractive motor-driven airplanes, the fifth was a rather dull glider. After the descriptions, the experimenter took a public census in which each of the boys announced which of the planes he favoured. Almost all the boys chose one of the motor-driven models. One member of each group, however, was a stooge who had been instructed to choose the glider. Since differences of opinion existed, the experimenter suggested that the boys discuss the problem and decide which of the planes they should build for the competition. The discussion was interrupted twice by the experimenter at seven and fourteen minutes for a quick census, and, after twenty minutes, was terminated by a last census. During the entire discussion the stooge maintained his opinion that the group should build the glider.

Following the discussion, the experimenter, under the pretext of getting the club organized, administered a number of sociometric measures and brief questionnaires.

With minor variations, the experiment, as described to this point, was identical in each of the seven participating countries. Following the administration of the measures, the procedure varied from country to country, as the experimental teams of each nation explored interests specific to the nation. Some countries were interested in the effect of various patterns of subgroup formation on "forgiveness" of the deviate stooge; others in determinants of productivity, etc. In this paper only the results of the common core experiment are reported.

Design

There were four experimental conditions.

(a) High Valence; Low Probability of reaching goal. (Abbreviated Hi Lo)
(b) High Valence; High Probability of reaching goal. (Hi Hi)
(c) Low Valence; Low Probability of reaching goal. (Lo Lo)
(d) Low Valence; High Probability of reaching goal. (Lo Hi)

Though originally it was planned to have an equal number of groups in each cell and a minimum of eight groups in any one condition, a variety of nation-specific difficulties made this impossible to achieve in all countries. The N in each cell in each country is presented in *Table I*. The figures in each

TABLE I *THE NUMBER OF GROUPS IN EACH CONDITION*

	Hi Lo	*Hi Hi*	*Lo Lo*	*Lo Hi*	*Additional groups*
Holland	10	10	10	10	13
Sweden	7	6	9	7	15
France	8	10	8	9	2
Norway	8	8	8	8	8
Belgium	6	6	6	6	14
Germany	8	8	8	8	24
England	4	2	2	3	19

column under the condition heading represent the number of groups in which the stooge was truly a deviate; that is, groups in which all boys except the deviate ended their discussion with unanimity of opinion as to which plane was the best to build. Under the column heading "Additional Groups" is listed the number of groups in each country in which the stooge was not a deviate. In such groups, either the subjects had not reached agreement among themselves or they were so enamoured of the glider that they eventually unanimously agreed with the stooge in his choice. Since the main object of this paper is the study of rejection of deviates and influence exerted on deviates, it will be concerned chiefly with those groups in which the stooge is a true deviate. Unless otherwise specified, all tables will present the data for only these groups.

Production of the Experimental Conditions

Valence is manipulated by varying the attractiveness of the goals. In high valence conditions, continuation of the club, and therefore the entire constellation of attractive ends which the group mediates, is at stake. In low valence conditions a minor and not particularly attractive sub-goal (free tickets to a dull documentary movie) is at stake.

1. High Valence—It is assumed that the demonstrations during the glamour period, and the promise of such things as meeting pilots, airplane rides, free movies, etc., created a set of extremely attractive goals for the boys. The possibility that their group would be discontinued and none of these goals would be attainable, may be counted therefore, an extremely unpleasant prospect. This possibility was introduced in this fashion:

> "I think now we had better talk seriously about all of the clubs that are being formed by the Royal Institute of Aviation. You probably remember that when I talked to your class in school, many of the boys in your class wanted to join these clubs. And much to our surprise this was true in all other classes and in all of the other schools we talked to. This means, of course, that a great number of boys want to join these clubs. Now we just hadn't expected this, we thought that there would be some interest, but we never guessed that there would be quite so much. And the situation now is this—in order to be completely fair, we felt that we must ask all of the boys who signed those sheets to one of these clubs. This will mean that we will have about twenty-five clubs. But the fact is that we simply do not have enough money or personnel or equipment to be able to keep all of those clubs going and it will be necessary to drop some of these clubs."

2. Low Valence is manipulated by creating a goal of relatively low valence or desirability. Thus:

> "The Institute has been able to get tickets to a showing of a documentary film on cargo planes. These are the planes that pick up and deliver freight. The film will show you what each member of the crew does, how they go about their work, technical details and information on where cargo is stowed and so on. Now we'd like to be able to give tickets to this interesting film to the members of all of the clubs but unfortunately we just do not have enough to go around and it looks as if some of the clubs just will not be able to get tickets for this movie."

Probability is manipulated by varying the proportion of potentially successful clubs to the total number of clubs, thus:

3. Low Probability of reaching goal conditions.

> "And I'm afraid that we'll only be able to keep five clubs going. This means, of course, that most of the clubs will have to be dropped. Of the total of twenty-five clubs, twenty will have to be dropped and only five can continue. You can see then, that any one club has very little chance to be chosen. And we can say that your club, like any other club, is very unlikely to be chosen to continue and actually has only a *very* small chance." (In low valence conditions substitute "getting movie tickets" for being chosen to continue.)

4. High Probability of reaching goal conditions.

> "But it really isn't too bad for we'll be able to keep at least twenty-two clubs going, very possibly all of them. This means, of course, that very few of the clubs will have to be dropped. Of the total of twenty-five clubs, at most two or three may be dropped and all of the others will continue. And, as I said before, though it may be necessary to drop two or three clubs, it may be possible to arrange things so that they all can continue. You can see, then, that you have a very good chance to be selected, your chances are very good and it is *very* likely that your club will be selected to continue."

In both high and low probability variations the experimeter employed a set of charts to dramatize the probability instructions.

Creating a Situation where Uniformity is Desirable for Locomotion

Before describing the five planes, and again immediately before the discussion of which plane to build, the experimenter specified the following ground rules for the competition: the boys could build only one model airplane and they must build this together; they must decide during the discussion which of the several planes they would build. Though all discussions lasted just twenty minutes, the experimenter said nothing to the boys about how long the discussion would last. If, during the course of the discussion, a boy attempted to convince the experimenter that a majority of the group wanted to build a particular plane and, therefore, that they should end the discussion and begin building, the experimenter non-committally continued the discussion with:

> "There still seem to be some differences among you, why don't you continue a while."

Under such conditions uniformity would facilitate locomotion, for the boys were unable to take even a first step toward their goal until they had decided upon a plane, and the discussion and argument caused by lack of agreement was clearly incompatible with working toward the goal.

The Stooge Deviate

The stooge was a schoolboy of the same age as the subjects and similar to them in appearance and manner. He chose the least attractive of the airplanes—the glider—and, in a standardized fashion, defended his choice until the experimenter ended the discussion. Since the boys were informed that they could decide upon only one model and must build this together, the stooge represented a barrier to group locomotion for 1. he championed a model that the other boys did not care for and did not think could win the competition 2. arguing with the stooge consumed much time and prevented the boys starting work on any model.

The stooge's behavior during the discussion was standardized and governed by the following conditions.

1. He was provided with three standard arguments with which to defend his position and every statement he made had to be a rephrasing of one of the standard arguments.
2. He spoke only when spoken to or when the position he was defending was under direct attack.
3. He had to speak at least once every four minutes. If during any four minute interval no remarks were addressed to him, he initiated a comment which was a rephrasing of one of his standard arguments.
4. The tone and manner of his remarks were objective and non-aggressive.

In each country, two boys alternated in the role of stooge. They appeared an equal number of times in each condition.[2]

Measurements

Two kinds of data were collected:

1. Post-discussion measurements of rejection, the effectiveness of the manipulated variables and of cohesiveness.
2. Observational data of the discussion of which airplane to build.

The post discussion measures were fitted into the club context of the experiment. Immediately after the final census, the experimenter said:

"That gives us a pretty good idea of how you feel. Now there are a few things I'd like to talk with you about. This is really your club and I'm only here to help you to get started and give you technical information and things like that. But you are going to have to think about how you would like to organize this club. You're going to have to decide such things as when you want to meet, how long each meeting should be, who you'd like to work with, who should be president, and so on."

The following measures were administered.

(a) A sociometric social preference scale:—On the pretext that work groups should be formed on the basis of the boys' personal preferences, each boy was asked to rate every other boy in the group on the following scale:

1. I would very much like to work with X.
2. I would like to work with X.

2. This is strictly true only of Holland, Norway, Belgium and Germany, where there were an equal number of groups in each cell. In the three other countries there was an unequal N in the four cells. Where N is even, the two boys appeared the same number of times. Where N is odd, one of the boys appeared just one time more than the other. Close examination of the differences between the two boys in the experiments concerned revealed that in no case can this slight imperfection of design account for differences between conditions.

3. I don't care whether or not I work with X.
4. I prefer not working with X.
5. I definitely do not want to work with X.

(b) Presidential ballot:—In order to choose a president for the club, each boy was requested to rank order all of the other boys in the order in which he felt they were most suitable for this important job. This ballot provides a measure of sociometric role preference.

(c) A cohesiveness questionnaire consisting of two questions:

1. Do you want to remain a member of this club?
2. How often would you like this club to meet?

(d) A measure of the success of the probability manipulation—to be described in the following section.

(e) A measure of the success of the valence manipulation—to be described in the following section.

Two observers were present at each experimental session. During the glamour period they helped with the demonstrations, played with the boys, and so forth, and they became an accepted and natural part of the club setting. During the discussion of which plane to build, one observer recorded who spoke to whom, and coded each remark according to the following set of categories:

1. Friendly: affectionate, pleasant comments and attempts to influence.
2. Objective: "plane" oriented comments; exposition, explanation or statements of own opinion.
3. (a) Imperatives: orders or strong suggestions to comply.
   (b) Majority arguments: attempts to exert pressure by pointing out that a boy is different or a minority group member.
4. Aggressive: angry, aggressive arguments; "objective" remarks voiced in loud, emotional tone.
5. Threatening: remarks threatening or indicating rejection.
6. Irrelevant: comments having no connection with the choice of a plane or with the attempt to influence.

A second observer employed a schedule designed to give indications of the "atmosphere" of the meeting and of the effectiveness of some of the manipulations. There were eight categories in this schedule—1. interruptions; 2. pauses; 3. out-of-field behavior; 4. side conversations; 5. chaos; 6. laughter; 7. dependence on leader; 8. time pressure. In addition the experimenter club-leader recorded changes of opinion during the discussion.

Follow-up

The problem of what to do with the boys after the first experimental meeting was handled differently in the various countries. Some of the

countries were fortunate enough to be able to run their clubs in the name of some local aviation group which was interested in increasing membership. After the first meeting, the boys automatically became members of such groups. The remaining countries were able to make arrangements with local air lines and aviation groups to treat the boys to such things as movies, visits to airports, and airplane flights, before the summer vacation mercifully set in and the clubs could be disbanded.

The Effectiveness of the Manipulations

A. Probability. Experience in the trial runs of the experiment indicated major difficulties in attempting to measure the success of the probability manipulation by an ordinary questionnaire. Many boys, for apparently superstitious reasons, consistently refused to put on paper an opinion that their club might not win the competition, despite the fact that in post experiment personal interviews they freely admitted that they had understood the experimenter's talk. In order to put a premium on realism the probability test questions were put into the framework of a forecasting contest. It was announced that the boy who could most accurately predict how well his club would do in the over-all competition would be awarded a model airplane kit complete with a working gasoline motor. The boys then answered the following questions.

(a) "Do you think that your club will be one of those selected to continue?" (or, "to receive movie tickets?")
—Yes. —No.

(b) "What chance do you think there is that your club will get movie tickets?" (or, "be selected to continue?")
1. . . . There is a very bad chance.
2. . . . There is a bad chance.
3. . . . There is a fifty-fifty chance.
4. . . . There is a good chance.
5. . . . There is a very good chance.

(c) On a list with twenty-five numbered positions, the boys indicated in which place among the total twenty-five clubs their own club would finish. This device was inserted simply to make the whole business seem somewhat more plausible. Data from this question cannot, of course, be used, for results are, in large part, an artifact of the fact that twenty-two clubs will win in high probability and only five in low probability. Thus, a boy in a high probability condition who thinks his club will win has twenty-two positions in which to place his club, whereas a boy in a low probability condition has only five such positions.

The answers to questions (a) and (b) are very highly correlated. Because

question (b) is a five point scale, and therefore considerably more useful for our purposes, we will rely primarily on the results of this question.

Results are presented in *Table II*. Each figure represents the mean for

TABLE II *THE EFFECTIVENESS OF THE PROBABILITY MANIPULATION*

	Hi Lo	*Hi Hi*	*Lo Lo*	*Lo Hi*
Holland	2·90	3·81	3·06	3·70
Sweden	2·91	3·80	2·83	3·78
France	2·87	3·88	2·92	3·66
Norway	2·73	4·09	2·59	4·11
Belgium	2·99	3·43	2·90	3·65
Germany	3·20	3·56	3·43	3·73
England	3·29	3·59	3·39	3·75

all groups in a condition of answers to question (b). The higher the figure, the better the estimated chances of being one of the winners. In *Table III*, are presented the results of the tests of significance of difference between conditions.[3]

TABLE III *RESULTS OF SIGNIFICANCE TESTS ON THE PROBABILITY MANIPULATION*

	Hi Lo vs. *Hi Hi*	*Lo Lo* vs. *Lo Hi*	*Hi Lo* + *Lo Lo* vs. *Hi Hi* + *Lo Hi*
Holland	·001	·002	·001
Sweden	·01	·001	·001
France	·001	·001	·001
Norway	·001	·001	·001
Belgium	·06	·002	·001
Germany	·08	·06	·02
England	·80	·66	·26

With the exception of England, between-condition differences are significant in all countries. It should be noticed, however, that there is considerable variation among the various experiments in the position of their scores along the probability scale and in the magnitude of difference between high and low probability scores.

B. Valence. The effectiveness of the valence manipulation was measured by answers to the question: "How disappointed would you be if your group

3. We have in this paper used a variety of statistical techniques. Most frequently the Mann-Whitney "U" test (7) has been used. Where the number of ties vitiated the value of this test the median test (10) or Pitman's Randomization tests were employed (8). In addition, we have used chi-square, the exact test for 2 × 2 tables and where the assumptions seemed satisfied, the t test. All levels of significance reported are for the two tails of the distribution.

was not one of those chosen to continue?" (or, "did not receive movie tickets?")

1. . . . I can't imagine anything that would disappoint me more.
2. . . . I would be extremely disappointed and would feel very badly.
3. . . . I would be disappointed and would feel badly.
4. . . . I would be a little disappointed, but it wouldn't make too much difference to me.
5. . . . I would not be at all disappointed.

TABLE IV *THE EFFECTIVENESS OF THE VALENCE MANIPULATION*

	Hi Lo	*Hi Hi*	*Lo Lo*	*Lo Hi*
Holland	2·06	2·00	3·00	2·79
Sweden	1·94	1·99	3·52	3·40
France	2·37	2·08	3·57	3·38
Norway	1·74	1·99	3·40	3·21
Belgium	1·84	2·23	2·74	2·83
Germany	2·43	2·24	3·67	3·10
England	1·85	1·99	2·15	2·56

Results are presented in *Table IV*. Each figure is the mean for all groups in a condition. The greater the figure, the less the disappointment and, by inference, the smaller the valence. *Table V* presents the results of the tests of

TABLE V *RESULTS OF SIGNIFICANCE TESTS ON VALENCE MANIPULATION*

	Hi Lo vs. *Lo Lo*	*Hi Hi* vs. *Lo Hi*	*Hi Lo* + *Hi Hi* vs. *Lo Lo* + *Lo Hi*
Holland	·004	·001	·001
Sweden	·001	·001	·001
France	·004	·002	·001
Norway	·001	·002	·001
Belgium	·01	·17	·01
Germany	·001	·008	·001
England	·80	·88	·62

significance. With the exception of England, the manipulation has been effective in all countries. Again, attention is drawn to the considerable variation among the various experiments.

METHODOLOGICAL AND PROCEDURAL STEPS SPECIFIC TO THE CROSS-CULTURAL ASPECTS OF THE STUDY

The decision to run similar experiments in all countries imposed on the over-all design features which were intended to insure some degree of

identity among the several experiments. Additionally, specific measures were introduced which would allow evaluation of the comparability of the several studies. In this section are described the features of the design specific to the cross-cultural nature of the project. In the following section the comparability of the several experiments will be evaluated.

Face similarity of the several experiments demanded at a minimum that:

1. the experimenters and stooges in the several countries perform in the same way and say essentially the same things;
2. the observers in the several countries have common understanding of the categories in the observation schedules;
3. the several measuring instruments be similar in wording.

A summary of the working procedures of the group is necessary to illustrate the procedures by which the several experiments were standardized.

The hypotheses to be tested and the variables to be manipulated and measured were agreed upon in two group meetings conducted in the Fall of 1951 and the Spring of 1952. Tentative decisions were also made about some of the specifics of a design that seemed feasible in the several countries. The Dutch team working with the research coordinator was assigned the job of running trial groups, to try out the design and work out the details of procedure. Following these trial runs, similar trial experiments were run in each of the participating countries. On the basis of this joint experience a manual was prepared which contained highly detailed, step-by-step descriptions of the procedure to be followed by experimenter and stooge in each phase of the experiment. Observers, too, receive a manual containing definitions of each category, examples, and pre-coded protocol as training material. Each country tried out several experiments on the bases of this manual, and in the Fall of 1952 a meeting of all group members was held to finalize the design of the experiment. The research coordinator then briefly assumed the role of travelling parameter and, with the assistance of translators, observed one experiment in each of the experimental conditions in each country. He attempted during this time to standardize the procedure in each country and eliminate variation from the agreed design. Following the observation of each experiment, the experimenter, stooge, coordinator, and translator worked together in role, playing the experiment, concentrating, of course, on those portions of the experiment that varied from the central design. In addition, the coordinator worked with each of the observers on problems of categorization and interpretation.

In order to ensure that the several measures employed in all experiments were equivalent, the following procedure was adopted. The originals of each instrument had been formulated in English, the working language of the group, then translated by each group member into his native tongue. In each country these translated instruments were then retranslated into English by two bilingual persons who had no knowledge of the experiment. The

retranslations and the English originals were compared and any differences eliminated. By this procedure it was possible to correct obvious misunderstandings, and to ascertain that the several versions of each measure were, at least, roughly equivalent. It was not possible, of course, to be sure that the distance between the points of the rating scales or the location of each point on the scale were identical in the several versions of each instrument.

To evaluate the extent to which the final version of the several experiments were comparable both in face similarity and conceptual equivalence, the following measures and procedures were instituted:

1. The comparability and success of manipulations of the independent variables were determined by the independent measures described earlier.
2. The comparative success of the attempt to create a situation where uniformity is perceived as facilitating locomotion was determined by observation of the frequency of subject-initiated comments indicating awareness of such a situation.
3. The comparability of the verbal behavior of the several experimenters and stooges was evaluated from two sources of data:
   (a) Observation of the frequency of their participation in the group discussion and of the "kind" of comments they made.
   (b) Verbatim transcriptions of the entire meeting of two groups in each country.
4. To evaluate the extent to which the observers in each country had common understanding of the categories in the observation schedules, a written protocol of two complete discussions was distributed to all observers in all countries during the course of the experiment. The observers coded each statement in the protocol in the categories of the observation schedules. The protocol was in English, a language which all the observers could read with facility. We shall not in this paper report in detail on the results of this "observer comparability check" but will refer to it only when necessary for the interpretation of results.

THE COMPARABILITY OF THE EXPERIMENTS

To test thoroughly the hypotheses under consideration, it is necessary that an experiment satisfy the following minimum criteria.

1. The manipulation of the independent variables must be effective.
2. The basic situation of the experiment must be completely apparent to the subjects; that is, the boys must perceive that uniformity is desirable in order to reach the goal and they must recognise the stooge as a deviate who blocks locomotion to the goal.

The results of the various experiments and the differences among them must, of course, be interpreted in terms of their success in satisfying these criteria. In this section the several experiments will be examined from this point of view.

1. The Manipulated Variables

(a) *Probability:* Ideally, the boys in the low probability conditions should have the impression that their chances are poor; in high probability conditions they should believe that their chances are good. The distance between these points on the probability dimension should be as great as possible.

Table II presents the probability measure data for only those groups in which the deviate is a true deviate. In *Table VI* the probability data for all

TABLE VI *EFFECTIVENESS OF THE PROBABILITY MANIPULATION*

1	2 *Low Prob.* *Hi Lo + Lo Lo*	3 *High Prob.* *Hi Hi + Lo Hi*	4 *Hi Prob. — Lo Prob.*
Holland	2·98	3·74	·76
Sweden	2·82	3·76	·94
France	2·91	3·77	·86
Norway	2·69	4·06	1·37
Belgium	3·02	3·68	·66
Germany	3·23	3·67	·44
England	3·10	3·70	·60

groups is presented. In some countries, groups that do not reach agreement among themselves, or that agree with the stooge, compose a major share of the total number of groups run (see column labelled "additional groups" in *Table I*). For this reason, the probability data in *Table VI* may be considered a somewhat more reliable index of the effectiveness of the manipulation than the data in *Table II*. The data of both low probability conditions is combined in the second column, and of both high probability conditions in the third column. In the fourth column are the differences between the two probability conditions.

It should be noticed in the low probability conditions that the mean scores in Belgium, Germany, and England are greater than 3·00 and tend to the "good chance" end of the scale. Despite the experimenter's instructions to the contrary, the boys in these countries tend to think they have a good chance of succeeding. These same three experiments also have the smallest scores in the high probability conditions. It follows, of course, that the differences between high and low probability conditions are smallest in the English, Belgian, and German experiments.

(b) *Valence:* In *Table VII* are presented the data on the valence manipulation. The figures recorded are the means of the "disappointment scale" for all groups run. Scores tending toward points 1 and 2 indicate extreme

TABLE VII *EFFECTIVENESS OF THE VALENCE MANIPULATION*

	High Valence Hi Lo + Hi Hi	*Low Valence* Lo Lo + Lo Hi	Lo Val. — Hi Val.
Holland	2·03	2·96	·93
Sweden	2·04	3·50	1·46
France	2·22	3·44	1·22
Norway	1·92	3·28	1·36
Belgium	2·02	2·81	·79
Germany	2·29	3·31	1·02
England	1·93	2·65	·72

disappointment and high valence; scores tending toward points 4 and 5 indicate relative indifference and low goal valence. It will be noticed in the low valence condition, that the means of the English, Belgian, and Dutch experiments are all below 3, an indication that the low valence manipulation has not been particularly effective in these countries.

II. Awareness that Uniformity Facilitates Locomotion to the Goal

As an index of how much the experimenter succeeded in creating among his subjects the perception that uniformity would facilitate locomotion to the goal, the researchers employed the frequency of tallies in the observation category "Time Pressure". This is defined as comments indicating that the subject is aware that the arguing and discussing caused by lack of uniformity within the group is incompatible with working toward the goal. Comments such as " If we have to keep arguing with you we'll never be able to get to work" or "We'll never win, if we keep on arguing forever" are included in this category.[4]

The frequency of such comments is recorded in *Table VIII.* In the first column are the mean number of such remarks in groups in which the subjects agree and the stooge is a true deviate. In such groups, Belgium and England have considerably fewer such remarks than any of the other countries. Germany too, is relatively quite low. In the second column are the mean number of "time pressure" comments, in groups in which the subjects do not agree among themselves. Again England, Belgium, and Germany are considerably lower than any of the remaining countries.

It is of interest to notice that in all countries the number of these comments is lower in the "subject-disagreement" groups than in the "subject-agreement" groups. This may be explained by the obvious fact that only in the "subject-agreement" groups is there a single deviate upon whom the

4. On the observer comparability check described earlier, the average intra-nation per cent agreement on precisely which statements are coded in this category is 77 per cent. The average inter-nation agreement is 66 per cent. *Rho* between the number of statements coded in this category in the actual experiments of each nation and in the written protocol is + ·07. It would appear, then, that there is good common understanding of this category among all of the observers in all of the countries, and the results of *Table VIII* cannot be accounted for as an artifact of differences between the observers in each country.

TABLE VIII "*TIME PRESSURE*" COMMENTS

	1 "*T.P.*" *Comments when s's agree*	2 "*T.P.*" *Comments when s's disagree*	3 *Per cent decrease* $\frac{\text{col. 1–2}}{\text{col. 1}}$
Holland	5·63	4·57	19
Sweden	3·24	2·47	24
France	4·17	4·00	4
Norway	4·25	3·71	15
Belgium	·92	·23	75
Germany	2·63	1·37	48
England	1·42	·43	70

group members can center their attention. Comments fitting the "time pressure" category are probably more likely to occur in such situations than when there are several dissenters. The degree of difference between the two kinds of groups is, however, particularly revealing of differences among the several experiments. In the third column of *Table VIII* is recorded the percent difference in the two kinds of groups. The difference is quite small in Holland, Sweden, France, and Norway, Sweden's 24 per cent decline being the largest figure in this group. The differences are all considerably larger in Belgium, Germany, and England. It seems reasonable to conclude from these data that "time pressure" comments in England, Belgium, and Germany are in large part prompted by the existence of a deviate, while in the remaining countries such comments occur whether or not there is a single deviate blocking locomotion. In the Belgian, German, and English experiments, there is an average of 65 per cent of all "subject-disagreement" groups in which not a single "time pressure" comment is recorded. In the remaining four countries, only 9 per cent of such groups have no "time pressure" comments.

One may question, further, whether these "time pressure" comments are indicative of the same events in Belgium, England, and Germany as in the other countries. Such comments could result from the fact that the subjects are aware of and concerned with the incompatibility between talking and working; in this case, they should be earnest, serious remarks. Or they could simply indicate amusement with a particularly stubborn boy, in which case one might expect they would frequently take the form of wisecracks. If such remarks are primarily wisecracks, high correlations should be anticipated between the number of "time pressure"[5] comments and the frequency of laughter. Relevant data follows:

5. The figures presented are all the mean rank order correlation within condition for "subject-agreement" groups. In England the correlations are calculated only in *Hi Lo* and *Lo Hi* conditions, for the other conditions have only two cases in a cell. It is impossible to compute such *rhos* in "subject-disagreement" groups because of the high number of English, Belgian, and German groups in which no "time pressure" comments are made.

	Mean *rho*
Holland	+·17
Sweden	+·06
France	+·22
Norway	+·33
Belgium	+·57
Germany	+·51
England	+·63

Clearly, the correlations are considerably higher in England, Belgium, and Germany than in any of the other countries. Not only, then, are there fewer "time pressure" comments in these three experiments, but there is some reason to believe that such comments in these experiments tend to indicate amusement at stubbornness rather than awareness that uniformity will facilitate locomotion. It seems reasonable to infer that the experiments of England, Belgium, and Germany have been less successful in creating the necessary experimental situation than have the Dutch, Nowegian, Swedish, and French experiments.

If our interpretation of "time pressure" comments in the several experiments is correct, it should follow that in those countries where "time pressure" comments are few, there should be a greater proportion of groups in which the subjects do not reach agreement among themselves.[6] If the subjects do not perceive that uniformity is desirable, pressures to uniformity should be low, and the subjects should be less inclined to change their own opinion and to attempt to influence the opinion of others. Pertinent data follows. Next to each country is reported the percent of the total number of groups run in which the subjects did not reach uniformity of opinion.

Holland	13%
Sweden	25%
France	6%
Norway	17%
Belgium	34%
Germany	32%
England	43%

It is clear that the English, Belgian, and German experiments have each a larger proportion of groups which did not agree among themselves than any of the other experiments. In these three countries an average of 36 per cent of the total number of groups run did not reach agreement; in the other experiments an average of only 15 per cent of the groups reached no agreement.

6. Groups in which the subjects end up by agreeing with the deviate's choice of the glider appear to be determined largely by the fact that in such groups the subjects tended initially to prefer the glider. In 80 per cent of such groups one or more subjects chose the glider in the pre-discussion choice of airplane. This was so in only some 20 per cent of the "subject-agreement" groups.

To summarize, on all of our major evaluative criteria the experiments of Norway, France, and Sweden seem clearly to establish the necessary conditions to provide a fair test of the hypotheses. Holland has a somewhat weak valence manipulation but satisfies all of the other criteria. The German experiment has a very weak probability manipulation and is quite weak in establishing the perception of uniformity as facilitating group locomotion. The English and Belgian experiments do not satisfy any of the three major criteria. It seems a fair inference, then, that the English, Belgian, and German experiments do not create the conditions necessary to test the hypotheses, and are not, therefore, comparable to the remaining experiments. Some of the possible reasons for this afford interesting illustrations of the difficulties of cross-cultural research where it is attempted to replicate precisely the same experimental design in all cultures, and provide illustrations for the proposition that phenomenal equivalence does not necessarily produce conceptual equivalence.

In Belgium, we have noted, the valence manipulation was relatively ineffective. In a supposedly low valence condition, the goal was apparently highly valent. The subjects in Belgium were parochial school boys in the Flemish part of Belgium, and subsequent inquiries revealed that for these boys movie going was a relatively rare privilege. A movie, then, any movie at all, was highly valent for these boys. Evidence supporting this suggestion is found in internal analysis of the Belgian disappointment scale data. Some of the schools from which the subjects were recruited had a more liberal "movie going" policy than others. Boys from the more liberal schools averaged 3·00 on the disappointment scale; the remaining boys averaged 2·66.[7]

In Germany the entire experimental set-up was far more difficult to carry off than in any of the other countries. At the time the experiment was conducted, Germany was still under control of the occupying forces. Until a few months before the experiment all activities connected with aviation had been prohibited. The prohibition was strictly enforced and extended to such minutiae as a ban on the sale of model airplanes.[8] The initial impact of recruiting for members for boys' aviation clubs was traumatic. A few of the teachers were violently opposed, parents were alarmed, and the Frankfurt police department launched an investigation to determine if this was an underground movement to revive militarism in Germany. Once explanations had been given, the official furore died down quickly, but parental

7. Similar difficulties presented themselves in planning the specific goal to coordinate to the high valence condition. At one time, goals such as a flight in an airplane or a de luxe all-day visit to an airport were considered for the high valence manipulation. Data gathered on the sign-up sheets which volunteers filled out for trial runs of the experiment indicated that 22 per cent of the volunteers in Sweden had flown and only 7 per cent of volunteers in Holland had flown. Similarly, 79 per cent of the boys in Sweden had visited an airport and only 43 per cent in Belgium had done so.

8. Data from the volunteer sheets mentioned in the previous footnote indicate that only 6·5 per cent of the German volunteers had built model airplanes. In the remaining 6 countries an average of over 32 per cent of the boys had done so.

anxiety remained high throughout the months during which the experiment was conducted. Though the German experiment was extremely successful at getting volunteers (91 per cent of all boys spoken to volunteered), it was quite unsuccessful at getting the boys to come to the first meeting (only 60 per cent of contacted volunteers came to the experiment). Almost invariably the explanation was, "My parents won't let me." The experimenters, too, were besieged by telephone calls from parents demanding some sort of explanation before giving their boys permission.

It need hardly be added that this atmosphere was not particularly propitious for running an experiment. The subjects were ill at ease and the experimenter tense. It would have been difficult to deliberately choose a more inappropriate experimental setting.

RESULTS

In this section are presented the major data on the dependent variables—rejection and communication. Though brief comments are given on certain major trends, and some of them are related to the results of the preceding section, detailed consideration of the similarities and differences in the results of the several experiments is reserved for later sections of this paper.

A. *Rejection*

In *Table IX* are presented the results of the sociometric working preference scale. The figures in each cell are derived for each group by subtracting

TABLE IX *REJECTION OF THE DEVIATE ON THE SOCIOMETRIC WORKING PREFERENCE SCALE*

	Hi Lo	*Hi Hi*	*Lo Lo*	*Lo Hi*
Holland	1·17	·65	·95	·70
Sweden	1·47	·95	1·28	·80
France	1·62	1·14	1·69	1·55
Norway	1·61	1·26	·85	1·36
Belgium	·82	1·09	1·08	·89
Germany	·56	·76	·32	·37
England	·71	·67	·65	1·01

the mean rating the subjects give to one another from the mean rating they assign to the deviate. Thus, if the subjects give the deviate a mean rating of 4 on the sociometric scale and give to one another a rating of 3, the rejection score is equal to 1·00. The larger the score the greater the rejection. The results of the test of significance are presented in *Table X*. Any difference that is significant at greater than the ·20 level of confidence is noted as n.s. On the left side of the tables are the results of the analysis of the probability variables; on the right the analysis of the valence variable.

TABLE X *RESULTS OF TESTS OF SIGNIFICANCE ON THE DIFFERENCES BETWEEN CONDITIONS ON THE SOCIOMETRIC SCALE SCORES*

	Probability Analysis			*Valence Analysis*		
	HL vs. *HH*	*LL* vs. *LH*	*HL*+*LL* vs. *HH*+*LH*	*HL* vs. *LL*	*HH* vs. *LH*	*HL*+*HH* vs. *LL*+*LH*
Holland	·03	n.s.	·02	n.s.	n.s.	n.s.
Sweden	n.s.	·20	·08	n.s.	n.s.	n.s.
France	·08	n.s.	·07	n.s.	·10	n.s.
Norway	·13	·15	n.s.	·08	n.s.	·15
Belgium	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Germany	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
England	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

The results of the presidental ballot are presented in *Table XI*. Each figure is derived by averaging the mean rank assigned to the deviate in each group.

TABLE XI *REJECTION OF THE DEVIATE ON THE PRESIDENTIAL BALLOT*

	Hi Lo	*Hi Hi*	*Lo Lo*	*Lo Hi*
Holland	4·93	4·67	4·65	4·34
Sweden	5·69	4·69	5·21	4·88
France	5·39	4·96	5·42	5·26
Norway	5·22	4·78	4·21	5·11
Belgium	4·42	5·08	4·70	4·37
Germany	4·18	4·35	3·75	4·04
England	4·14	3·90	2·91	4·49

The higher the figure the greater the rejection. Results of the tests of significance are presented in *Table XII*.

It can be seen that in several countries sampled rejection appears to be

TABLE XII *RESULTS OF SIGNIFICANCE TESTS ON THE PRESIDENTIAL BALLOT*

	Probability Analysis			*Valence Analysis*		
	Hi Lo vs. *Hi Hi*	*Lo Lo* vs. *Lo Hi*	*Hi Lo*+*Lo Lo* vs. *Hi Hi*+*Lo Hi*	*Hi Lo* vs. *Lo Lo*	*Hi Hi* vs. *Lo Hi*	*Hi Lo*+*Hi Hi* vs. *Lo Lo*+*Lo Hi*
Holland	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Sweden	·005	n.s.	·03	·08	n.s.	n.s.
France	·16	n.s.	·18	n.s.	n.s.	n.s.
Norway	·19	·05	n.s.	·04	n.s.	·18
Belgium	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Germany	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
England	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

virtually a universal reaction to the deviate. In all countries and in all conditions the deviate is considered relatively undesirable as either a working partner or a club president. The single exception is the *Lo Lo* condition in England, where on the presidential ballot the deviate is ranked better than the median rank of 3·5. This figure, however, is based on only two cases.

There is, in general, a high degree of similarity between the presidential ballot and the working-preference scale in the relative magnitudes of rejection in the four conditions. The only exceptions are the reversals in *Hi Hi* and *Lo Hi* in Sweden and Holland.[9]

On between-condition comparisons it will be noticed first that there are no significant or near significant differences on either measure in any of the condition comparisons in England, Belgium, or Germany. This, of course, is what one would expect when the experimental conditions necessary to test the hypotheses are not satisfied. In the remaining four countries, the following tendencies appear:

1. In Holland, Sweden, and France, probability would appear to be the determining variable. In these countries, the combined *Lo* probability scores are consistently greater than the combined *Hi* probability scores. In Norway, however, the combined scores would seem to indicate no effect of probability and a tendency for the combined *Hi* valence groups to reject more strongly than the *Lo* valence groups.
2. In all cases, rejection in *Hi Lo* is greater than in *Hi Hi*. Though the significance levels in some of the experiments allow us to speak only of tendencies, the combined probabilities[10] for either measure yield close to the ·01 level of confidence and the null hypothesis may be rejected.
3. There is a tendency in Holland, Sweden, and France for rejection in *Lo Lo* to be greater than in *Lo Hi*. Norway, however, has a distinct tendency in the opposite direction.
4. With the exception of France, rejection scores are somewhat greater in *Hi Lo* than in *Lo Lo*. There are slight indications of an opposite trend in the two high probability conditions.

B. *Communication*

The observer's records of the discussion of which plane to build serves

9. It is interesting, however, to note that there are differences among the countries in the effectiveness of the two measures. In Sweden, the presidential ballot yields far stronger statistical results; in Holland and France the working-preference measure does so. There is some reason to believe that boys of this age in Sweden are somewhat more accustomed to formal organization and activities than are similarly aged boys in Holland and France. Possibly, then, boys in Sweden are more accustomed to making judgments of role suitability, more sensitive to the implications of deviation in a leader, and treat this as a more important matter than the expression of preferences for obviously temporary working partners.

10. Probabilities combined using the formula $x^2 = -2 \log_e P$ for $d.f. = 8$. See Kendall, A. G.: *Advanced Theory of Statistics*, Vol. II (5).

as the source of data indicating the strength of the tendency to influence others. We shall be concerned chiefly with the attempt to influence the deviate.

The difficulties of achieving a proper index to coordinate to the notion "attempt to influence others" have been documented in previous studies of the influence process (9). In all cases, the major problem has been the fact that not all communications are attempts to influence. This was particularly true in the discussions of these young boys. Occasionally the discussion was completely irrelevant to the topic at hand but was prompted by some extraneous event such as a fire engine passing in the street or a boy falling from his seat. Sometimes discussion was simply an interested attempt to exchange information about airplanes. Remarks to the deviate were sometimes made only because a boy was prodded by his neighbours into making such a remark, and so on.

The observation schedule described earlier is an attempt to discriminate among such communications. It is assumed that comments coded in the categories "imperative or minority arguments", "aggressive", and "threatening" may unequivocally be classified as attempts to influence. Presumably, the subjects will initiate such "intense" communications only when they are seriously attempting to influence one another. Comments coded "irrelevant" would clearly not indicate an attempt to influence and would be stimulated largely by some external event. Comments coded "friendly" or "objective" may be attempts to influence or may represent simply an exchange of information or a spontaneous joke. As an index of the strength of the tendency to influence others the following ratio is used:

$$\frac{\textit{No. of "imperative-minority", "aggressive" and "threatening" remarks to the deviate}}{\textit{Total No. of subject initiated comments—No. of Irrelevant comments}}$$

Our index, then, is the proportion of all relevant comments initiated by the subjects which are "intense" communications to the deviate.[11]

The amount of communication to the deviate will, of course, be dependent upon the time when the subjects reach complete agreement among themselves, at which point the stooge becomes a true deviate. This time varied somewhat from group to group. In order to make the communication data of all groups comparable, it is necessary to choose a period for analysis following the time the subjects reached agreement. The seven minute between-census period immediately following the time that the subjects achieved consensus is the interval on which the analysis of the communication data is based. This means, of course, that those groups which reached

11. Reliability of categorization of comments to the deviate was computed by dividing the protocol of each observer into minute intervals, tallying in each interval separately the number of agreements, and computing for the twenty minute discussion the percent agreement between observers. When the six categories are combined into three: "friendly" + "objective"; "imperative" + "agressive" + "threatening", and "irrelevant", there is an average of 65 per cent agreement in the several countries. The uncombined categories yield an average of 63 per cent agreement.

TABLE XIII *NUMBER OF CASES ON WHICH THE COMMUNICATION ANALYSES IS BASED*

	Hi Lo	*Hi Hi*	*Lo Lo*	*Lo Hi*
Holland	9	8	8	8
Sweden	5	6	9	4
France	7	9	8	7
Norway	6	6	8	8
Belgium	3	6	6	3
Germany	4	7	5	5
England	2	2	1	4

agreement only after the census at the fourteen minute mark are eliminated from the analysis. The number of groups in each country on which the communication analysis is based is presented in *Table XIII*.[12]

TABLE XIV *COMMUNICATION TO THE DEVIATE*

	Hi Lo	*Hi Hi*	*Lo Lo*	*Lo Hi*
Holland	·24	·23	·27	·21
Sweden	·16	·36	·24	·17
France	·09	·18	·23	·24
Norway	·23	·42	·24	·22
Belgium	·05	·04	·03	·05
Germany	·05	·13	·21	·18
England	·07	·06	·02	·04

The results of the communication analysis are presented in *Table XIV*. Significance levels of the difference between conditions are presented in

TABLE XV *RESULTS OF SIGNIFICANCE TEST ON COMMUNICATION TO DEVIATE*

	Hi Lo vs. *Hi Hi*	*Lo Lo* vs. *Lo Hi*	*Hi Lo* vs. *Lo Lo*	*Hi Hi* vs. *Lo Hi*
Holland	n.s.	n.s.	n.s.	n.s.
Sweden	·05	n.s.	n.s.	·11
France	·18	n.s.	·04	n.s.
Norway	·18	n.s.	n.s.	·10
Belgium	n.s.	n.s.	n.s.	n.s.
Germany	n.s.	n.s.	n.s.	n.s.
England	n.s.	n.s.	n.s.	n.s.

12. Additionally, groups are eliminated which make a total of ten or fewer relevant comments during the seven minute period of analysis. There are five such groups—one apiece in Holland, Belgium, and Germany and two in Norway. In such groups the boys were so extremely shy that literally almost nothing was said, or so absorbed by some sudden outside event that for this period there was almost no discussion of the topic. For example, in one Dutch group, the boys, highly excited by an outside happening, made the amazing total of 118 comments in seven minutes of which 110 were irrelevant. In such groups nothing is at test but the experimenter's patience.

Table XV. Again it will be noted that none of the differences in Belgium, Germany, and England approach even the ·20 level of confidence.[13] In the remaining countries, the following trends are noted.

1. Excepting the Dutch experiment, "intense" communication to the deviate is greater in *Hi Hi* than in *Hi Lo*. In the two low valence conditions there is a very slight trend in the opposite direction.
2. In all four countries, communication to the deviate is greater in the *Lo Lo* than in the *Hi Lo* condition. Excepting France, communication to the deviate is greater in *Hi Hi* than in *Lo Hi*.

It should be noted that with the exception of the *Lo Lo–Lo Hi* relationships, the major trends in the communication data are exactly opposite those in the rejection data.

INTERPRETATION

There is clear indication that rejection is a general and widespread reaction to a deviate. In all countries and in virtually all conditions the deviate is considered relatively undesirable. This finding is similar to that of several studies (3, 9) conducted in the United States where, again, deviates from a group norm were rejected. The nations involved in these several studies represent a sizeable sample of nations identified as members of "Western culture". For the large group of nations so identified, the demonstrated relationship between deviation and rejection would certainly appear to be a well established and safely generalizable finding.

Comparison among countries of between-condition results is complicated by the fact that the various experiments have been differentially successful. The data evaluating the several experiments has indicated that, for a variety of reasons, the experiments of England, Belgium, and Germany have not produced the conditions necessary to test the hypotheses. These experiments, then, are not comparable to the other four experiments, and attempts at interpretation must be limited to the data of the Norwegian, Swedish, Dutch, and French experiments.

For these experiments there are a number of possible alternative interpretations. Many of the statistical comparisons are non-significant, and the results of others yield extremely tentative levels of confidence. Therefore, on the basis of the condition comparisons which yield conventionally acceptable levels of significance (and the combined probability levels where the results of the several experiments are in the same direction), the following conclusions for the rejection variable could be drawn. If in a group locomotion situation, where the goal is threatened, a deviate blocks locomotion to

13. The Dutch data, too, reveals no difference between conditions, and the scores in each cell are extremely close. Possibly this is an observer artifact. The Dutch observers have the lowest reliability of all seven countries with only 51 percent agreement.

the goal, the probability of reaching the goal would appear to be the major determinant of the degree of rejection of the deviate. Though, in most of the experiments, there appears to be little indication that the valence of the goal effects rejection, the two variables appear to interact, for the effect of the probability variable is consistently at its greatest when the valence of the goal is high. Though such a conclusion seems plausible and defensible, there are sufficient disturbing exceptions to arouse doubts concerning the facility of the conclusion. For the moment, we shall be concerned with the most obvious exceptions to this conclusion, those in the Norwegian experiment, where we find:

1. No significant differences in the comparisons of the two high probability conditions against the two low probability conditions. Indeed on either rejection measure the mean of the high probability conditions is slightly greater than that of the low probability conditions.
2. A tentative tendency for valence to effect rejection.
3. Distinctly opposite results to those of the other countries in the *Lo Lo* vs. *Lo Hi* comparison. Whereas in the French, Swedish, and Dutch experiments rejection, on either measure, is greater for *Lo Lo* than *Lo Hi*, in Norway rejection is greater in *Lo Hi* than *Lo Lo* (significant on the presidential ballot at the ·05 level).

Such conflicting results leave us with a number of alternatives in interpreting our data.

(a) The Norwegian results are attributable to experimental artifacts.
(b) There is a cultural difference between Norway and the other three countries.
(c) The relationship among our variables is more complex than we originally thought and it may be possible to construct a theory which will reconcile these apparent differences.

The interpretation of the Norwegian results as attributable to experimental artifacts does not, from inspection of all the data, appear to be an easily acceptable explanation. We note, first, from *Table VII*, that the Norwegian valence manipulation was no more effective than that of Sweden or France. In *Table VI*, it is immediately apparent that the Norwegian experiment was by far the most successful of all experiments at manipulating the probability variable. It seems strange indeed that of all experiments the Norwegian one should show no effect of probability. Though certainly other factors which we have not measured could effect these results, close study of the verbatim protocol of the several experiments reveals nothing but the most trivial of verbal differences between the Norwegian experimenter and the experimenters of the other countries. In the data and records at hand it is impossible to identify any procedural artifacts which could account for these differences.

An interpretation involving cultural differences is a possible explanation of these results. For the finding that valence seems to be the determining variable in Norway and probability the determining variable in the other countries one could, for example, make the suggestion that Norwegians are strongly goal-oriented people, indifferent to the chances of success, ruthlessly sweeping aside all barriers if the goal is desirable, more tolerant when they are indifferent to the goal. The French, Dutch, and Swedish, on the other hand, might be described as more "success-oriented", more vulnerable to ego damage, and intolerant of deviates chiefly when they are threatened with probable failure, no matter whether the failure is serious or trivial. Hypotheses such as these, however, tend to be cognitively unsatisfying. Any number of such explanations could be devised and there is no strong reason arising from the data or knowledge of the cultures involved to prefer one explanation above the others. Not the least cause for uneasiness with this kind of explanation is the fact that any such culturally oriented hypotheses must in some way maintain cultural similarities between Sweden and France and cultural differences between Sweden and Norway.

Nevertheless, one cannot, at this point, discount any such explanation. The necessity for exploring more fully such explanations will depend on the eventual usefulness of our attempt to integrate these apparently divergent results into one conceptual scheme. Ideally, such a theory should reconcile not only the Norwegian exceptions but also the less marked deviations noted in the previous section. Additionally, an attempt will be made to incorporate the rejection and communication data in one formulation.

The theory presented concerns the relationship of three constructs to the set of dependent and independent variables. These constructs are 1. cohesiveness, 2. the tendency to redefine the group and 3. the tendency to integration.

1. *Cohesiveness* has been defined (3) as the resultant of all forces acting on members to remain in the group. It has been clearly demonstrated in a number of previous studies (1, 3, 9) that rejection and communication increase monotonically with increasing cohesiveness.

It is assumed that cohesiveness, within the limits of the manipulations, is independent of the valence and probability variables. Cohesiveness, then, should be roughly the same in all conditions. Data testing this assumption are presented in *Table XVI*. The figures in this table are derived from answers to the question, "How frequently do you think this group should meet?" Each figure represents the percentage of boys checking "once every 2, 3, or 4 weeks". These positions were at the low end of the provided scale; therefore the higher the figure the lower the cohesiveness. Except for France[14] the

14. The two low valence conditions in France are somewhat more cohesive than the high valence conditions (difference significant at ·03 level of confidence). The French experimenter, apparently embarrassed at the "lowness" of the low valence goal, reports that in describing this goal his manner was somewhat pleasanter than during his manipulation of high valence. This could, of course, account for the difference between these conditions.

TABLE XVI *THE RELATIONSHIP OF COHESIVENESS TO VALENCE AND PROBABILITY*

	Hi Lo	*Hi Hi*	*Lo Lo*	*Lo Hi*
Holland	·02	·00	·00	·04
Sweden	·08	·00	·08	·02
France	·10	·18	·05	·02
Norway	·04	·05	·02	·04

assumption seems to be justified. Within the remaining countries the figures are all fairly close to one another and there are no significant differences between any pair of conditions.

2. *The tendency to redefine the group* has been defined as the tendency to exclude from group membership those persons who deviate from group opinion. In previous studies (3, 9) it has been demonstrated that deviates are rejected. It is assumed that the strength of this tendency is a monotonically increasing function of (a) valence and (b) probability. Further, as the tendency to redefine the group increases, rejection (by definition) should increase and communication decrease. Evidence supporting the assumed communication decrease is found in studies by Schachter (9) and Festinger and Thibaut (4).

3. *The tendency to integrate* is defined as the tendency to restore harmony and reduce conflict between group members. It is assumed that such a tendency is active chiefly in group locomotion situations, where there is disagreement among the group members and where such disagreement threatens to disrupt progress toward the goal. It is assumed that the experimental situation my be characterized in this way; not only does the deviate block agreement on the best plane to choose, but conflict with the deviate could lead to less effective building of the plane and generally less efficient group functioning. As the tendency to integrate increases, rejection should decrease and communication to the deviate increase.

It is assumed that the tendency to integrate increases monotonically with increasing valence and decreasing probability. There is some evidence for this assumption. Lanzetta (6) in a study of stress (which may be considered as roughly analogous to our probability) reports "Under increased stress there was a decrease in behaviors which would tend to cause tension, friction, or disequilibrium in the group and an increase in behavior which would tend to cause a decrease in tension and increased integration of the group."

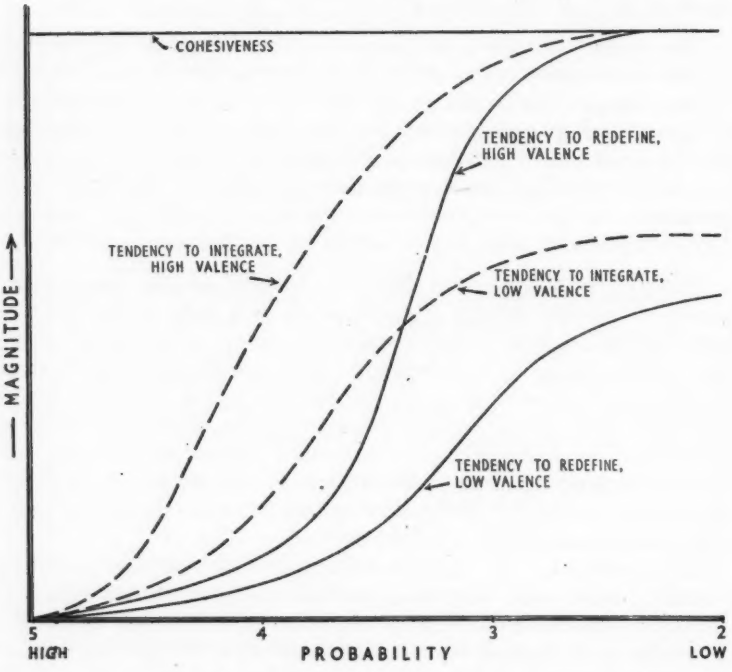
To summarize, we have assumed with some support from other studies, the following relationships between these three constructs and the independent variables.

1. Cohesiveness is independent of valence and probability.
2. The tendency to redefine the group increases monotonically with increasing valence and decreasing probability.

3. The tendency to integrate increases monotonically with increasing valence and decreasing probability.

These relationships are depicted graphically in *Figure* 1 where it will be noticed that cohesiveness is a straight line parallel to the horizontal axis, the high valence curves for both tendencies are higher than the low valence

FIGURE 1. *THEORETICAL CURVES RELATING COHESIVENESS, TENDENCY TO INTEGRATE, AND TENDENCY TO REDEFINE TO VALENCY AND PROBABILITY*

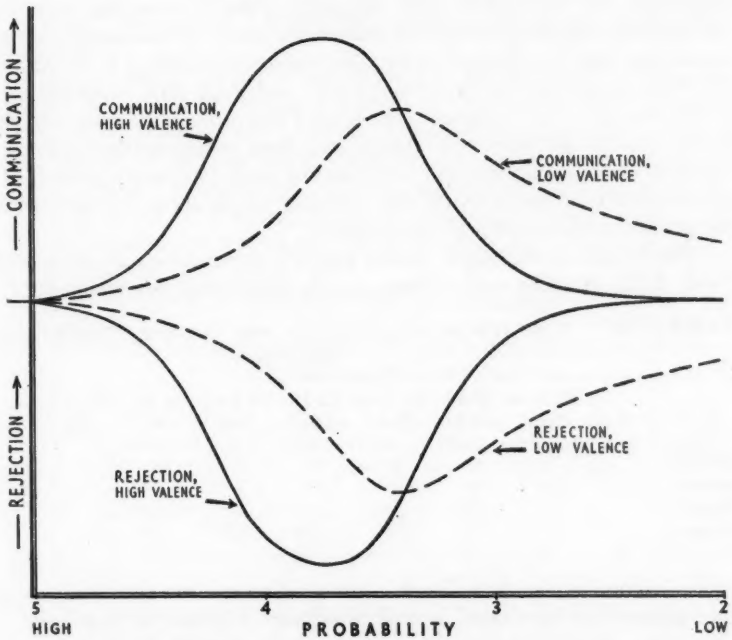


curves for these tendencies, and all tendency curves increase with decreasing probability. It should be further noticed that we make the following graphic assumptions—the high valence curves grow more rapidly and reach their asymptote sooner than the low valence curves. The precise height of the cohesiveness curve is, of course, arbitrary. It may be safely assumed, however, that in all countries, cohesiveness is very high. Of the total of almost 750 subjects in the four countries under consideration, not a single boy answered "No" to the question "Do you want to remain a member of this club?"

Between these constructs and our dependent variables we have made, and partially supported, the following assumptions.

1. Rejection and communication to the deviate vary directly with cohesiveness.
2. As the tendency to redefine the group increases, rejection of the deviate increases and communication to him decreases.

FIGURE 2. *DERIVED CURVES OF REJECTION AND COMMUNICATION*



3. As the tendency to integrate increases, rejection of the deviate decreases and communication increases.

More concisely stated:

Rejection = *cohesiveness* + *tend. redefine group* − *tend. integrate*
Communication = *cohesiveness* − *tend. redefine group* + *tend. integrate*

From these formulations it is possible, by simple algebraic manipulation

of the values of the curves in *Figure* 1, to derive a set of curves for the relationship of rejection and communication to the independent variables. These derived relationships are presented in *Figure* 2. The rejection curves are U-shaped with rejection highest when probability is very high or low. Communication curves approximate bell-shapes, with communication lowest when probability is very high or high. To the point that the two rejection curves intersect, rejection is higher, at a given point along the probability axes, in low valence than in high valence. Beyond the intersection point the reverse is true. These relationships are precisely the opposite for the two communication curves.

The comparison of experimental results in these curves is performed by projecting the probability scale values in *Table II* onto the appropriate curves, and comparing within any country the predicted relationship between conditions with the obtained relationship reported in *Table IX*, "rejection on the sociometric working preference scale", and *Table XIV* "communication to the deviate".[15] For example, in Holland the probability value for *Hi Lo* is 2·90 and for *Hi Hi* is 3·81. Comparing these points on the high valence rejection curve, it can be easily seen that the prediction is rejection in *Hi Lo* greater than rejection in *Hi Hi*. Our obtained results are in this direction and the predicted relationship is corroborated.

The results of the major within country comparisons are presented in *Table XVII*. Heading each column are the two conditions which are compared. Under the heading "Pred." (for prediction), the customary symbol for "greater" or "less than" is used to indicate whether the first listed condition in a column heading should, from the curves, be greater or less than the second listed condition. Thus, in the column headed "*Lo Lo* vs. *Lo Hi*" rejection should, by prediction, be greater in *Lo Lo* than in *Lo Hi* in France.

TABLE XVII *COMPARISON OF PREDICTED AND OBTAINED RESULTS ON REJECTION*

	Hi Lo vs. *Hi Hi*		*Lo Lo* vs. *Lo Hi*		*Hi Lo* vs. *Lo Lo*		*Hi Hi* vs. *Lo Hi*	
	Pred.	*Obt.*	*Pred.*	*Obt.*	*Pred.*	*Obt.*	*Pred.*	*Obt.*
Holland	>	+	≥	+	>	+	<	+
Sweden	>	+	≥	+	>	+	<	−
France	>	+	>	+	>	−	<	+
Norway	>	+	<	+	>	+	<	+

15. Ideally, of course, such comparisons should be made on the basis of a family of curves; that is, on the basis of the relative values of cohesiveness and valence in each country a set of curves should be drawn specific to each country. Such a procedure would allow us not only to study relationships within countries as we do now, but to compare countries. Unfortunately, such procedure would demand a high order of between country comparability on all of our major measures. At least one of our measures, however, is almost completely culture bound. Answers to the cohesiveness questions are determined, in good part, by such factors as the number of free hours the boys have, whether they are legally permitted to work after school hours, the freedom their parents permit them, etc. All of these matters vary considerably from country to country.

The reverse should be true in Norway. The symbol ≥ is employed where the prediction is a very small difference in the indicated direction. Under the heading "Obt." (for obtained) the symbol + is employed if the obtained results are in the same direction as the prediction; the symbol − is used if predicted and obtained differ.

In all but two cases, predictions and results agree. Of the two exceptions, the French *Hi Lo* vs. *Lo Lo* comparison may be explained by the cohesiveness differences between the high and low valent conditions. It will be remembered that cohesiveness is significantly greater in the French low valence condition, a fact which should increase the rejection in these conditions. This also, of course, helps explain the fact that in France the difference between *Lo Hi* and *Hi Hi* is considerably greater than in any of the other countries.

It should be noted that with this formulation the difference between the Norwegian *Lo Lo–Lo Hi* relationship and that of the other three countries is accounted for by means of a U-shaped relationship between rejection and probability plus the fact that the Norwegian high probability manipulation is so much "higher" than that of the remaining countries. By this formulation, then, it would be maintained that the Norwegian results are a manifestation of the same set of laws as seem to be operative in the experiments of Holland, Sweden, and France.

A similar analysis of the communication data is presented in *Table XVIII.*

TABLE XVIII *COMPARISON OF PREDICTED AND OBTAINED RESULTS ON COMMUNICATION TO DEVIATE*

	Hi Lo vs. *Hi Hi*		*Lo Lo* vs. *Lo Hi*		*Hi Lo* vs. *Lo Lo*		*Hi Hi* vs. *Lo Hi*	
	Pred.	*Obt.*	*Pred.*	*Obt.*	*Pred.*	*Obt.*	*Pred.*	*Obt.*
Holland	<	−	≥	+	<	+	>	+
Sweden	<	+	>	+	<	+	>	+
France	<	+	≦	+	<	+	>	−
Norway	<	+	>	+	<	+	>	+

It will be remembered that it was necessary to eliminate a number of groups in each country in making an analysis of the communication data. The comparison between predicted and obtained results, then, must be based on the probability scale values of only those groups on which the communication analysis is based. These values are presented in *Table XIX.* The communication data have been reported in *Table XIV.*

Prediction and obtained agree in all but two cases. Again, the exception in the French experiment may be explained by the differences in cohesiveness between high and low valent conditions. Cohesiveness in *Lo Hi* is considerably greater than in *Hi Hi*, a factor which, as has been demonstrated in several previous studies (1, 9) tends to increase communication. For the Dutch reversal, there would appear to be no apparent explanation, though

TABLE XIX *PROBABILITY SCALE VALUES FOR GROUPS ON WHICH THE COMMUNICATION ANALYSIS IS BASED*

	Hi Lo	*Hi Hi*	*Lo Lo*	*Lo Hi*
Holland	2·88	3·81	3·00	3·76
Sweden	2·77	3·80	2·83	3·96
France	2·82	3·89	2·92	3·71
Norway	2·69	4·01	2·59	4·11

the previously noted low reliability of observation in this experiment is a possible clue.

Though fourteen of sixteen predictions are confirmed and one of the two exceptions seems explainable, much of the communication analysis must be treated as tentative evidence. There are relatively few cases in some of the conditions, and reliability is unimpressive in some of the countries. In several of the comparisons, the differences are only a few percentage points. In such cases, agreement between prediction and result may be considered more a fortunate coincidence than a testimony to a theory.

It should be clearly pointed out that we do *not* consider this theory well substantiated at this point. There is no doubt that the suggested theory is presented in far more specific form than is warranted by the immediate data. We have deliberately chosen to formulate a theory in semi-precise form, rather than take advantage of the safety of a vaguer formulation, in order to spell out completely one scheme within which it may be possible to reconcile apparent divergent results. In its present form, this formulation fits the data reasonably well. Of the thirty-two between condition comparisons on two independently measured variables, there are only four cases in which predicted and obtained results disagree. Two of these exceptions seem explainable by variations in cohesiveness. The theory proposed has the additional virtues of integrating communication and rejection data into one conceptual scheme, and providing a framework within which it is possible to reconcile apparent differences between experiments into one set of lawful relationships. Pleasing though such virtues may be, however, they have little to do with the verification of the theory. That data and derivation generally agree is gratifying but scarcely surprising, for this is an *ad hoc* construction. The test of the usefulness of this formulation must be future experiments specifically designed to test its implications.

SUMMARY

An experiment is described which was simultaneously conducted in Holland, Sweden, France, Norway, Belgium, Germany, and England. The relationship between desirability of goal and probability of goal achievement to the tendencies to change others and to reject deviates is studied.

The several experiments differ in the extent to which they create the

conditions necessary to test the hypotheses. In three of the experiments there is good evidence that the manipulations of the independent variables have not been effective, and that the basic situation of the experiment was not clear to the subjects. The subjects did not perceive that uniformity was desirable in order to reach the goal. This failure to produce the necessary experimental conditions appears, in part, to be culturally determined. Though the experiments in all countries had face similarity, nation specific factors appear to have effected the extent to which particular instructions and manipulations were effective.

Rejection appears to be virtually a universal reaction to a deviate. In all countries and in all conditions the deviate is considered relatively undesirable on both sociometric and role preference measures. The effects of the probability and valence variables are studied in the four experiments in which the necessary experimental conditions have been successfully created. The experiments of Holland, Sweden, and France yield essentially similar results. In all three experiments:

1. The combined low probability condition rejection scores are consistently greater than the combined high probability scores.
2. Rejection is greater in *Hi Lo* than in *Hi Hi*.
3. There is a tendency for rejection in *Lo Lo* to be greater than in *Lo Hi*.

The results of the Norwegian experiment are on some points opposite to those of Holland, Sweden, and France. Valence, rather than probability, appears to determine the degree of rejection. Rejection in the *Lo Hi* condition is greater than in the *Lo Lo* condition.

Three possible explanations of these differences among experiments are considered.

1. The difference may be attributable to experimental artifacts.
2. The difference may be attributable to cultural factors.
3. The differences may reflect a more complex relationship than was originally supposed, and it may be possible to construct a theory which will reconcile these apparent differences. If the relationship between probability and rejection is U-shaped, it could be maintained that one set of laws can account for the results of all the experiments. A theory is suggested from which it is possible to derive such a relationship.

REFERENCES

1. BACK, K. "Influence through Social Communication." *J. Abn. & Soc. Psych.*, 1951, *46*, 9–23.
2. FESTINGER, L. "Informal Social Communication." *Psych. Rev.*, 1950, *57*, 271–282.
3. FESTINGER, L., SCHACHTER, S., and BACK, K. *Social Pressures in Informal Groups*. New York, Harper, 1950.

4. FESTINGER, L. and THIBAUT, J. "Interpersonal Communication in Small Groups." *J. Abn. & Soc. Psych.*, 1951, *46*, 92–99.
5. KENDALL, M. G. *The Advanced Theory of Statistics.* Vol. II. London, Griffin & Co.
6. LANZETTA, J. T. "Group Behavior under Stress." *Amer. Psychol.*, 1953, *8*, 383 (abstract).
7. MANN, H. B. and WHITNEY, D. R. "On a Test of whether One of Two Random Variables Is Stochastically Larger than the Other." *Ann. Math. Statist.*, 1947, *18*, 50–60.
8. PITMAN, E. J. G. "Significance Tests which May Be Applied to Samples from Any Populations." *J. Royal Statist. Soc. Suppl.*, 1937, *4*, 119–130.
9. SCHACHTER, S. "Deviation, Rejection and Communication". *J. Abn. & Soc. Psych.*, 1951, *46*, 190–207.
10. WESTENBERG, J. "Significance Test for Median and Interquartile Range in Samples from Continuous Populations of Any Form". *Proc. Kon. Ned. Akad. v. Wetensch.*, 1948, *51*, No. 2, 252–261.

BIOGRAPHICAL NOTES

STANLEY SCHACHTER is Associate Professor of Psychology and a member of the Laboratory for Research in Social Relations at the University of Minnesota.

JOSEF NUTTIN is Professor of Psychology and Social Psychology at the University of Louvain (Belgium). His major research interests are in the areas of human motivation, dynamic, and cognitive factors in personality and group behavior. His publications include *Psychoanalysis and Personality: A Dynamic Theory of Normal Personality*, Sheed and Ward, 1950; *Tâche, réussite et échec: théorie de la conduite humaine*, Louvain, Publications Universitaires, 1954; and, *The Place of Social Psychology in the Social Sciences*, Unesco publication, 1954.

PAUL H. MAUCORPS is *chargé de recherches* at the Centre National de la Recherche Scientifique and directs a team of social psychologists at the Centre d'Études Sociologiques in Paris. His team is primarily concerned with group experiments and surveys relating to the social attitudes of children and adolescents. He was formerly a Senior Psychologist in the French Air Force and editor of the monthly review *Travail et Méthodes*. His publications have included *Le Service d'Orientation et de Sélection du Personnel de l'Armée de l'Air*, Paris, 1947; *Psychologie militaire*, Paris, 1948; and *Psychologie des mouvements sociaux*, Paris, 1950.

DIEDRICH OSMER began his studies at the universities of Berlin and Munich in Economics, Sociology, and Law. In 1950 he joined the staff of the re-established Institute of Social Research at the University of Frankfurt where

under the directorship of Professors Max Horkheimer and Theodor W. Adorno he participated in the empirical studies conducted by this institute. He received his doctor's degree in 1953 with a dissertation on "The Group Discussion Method".

HUBERTUS C. J. DUIJKER studied philosophy and psychology at the University of Amsterdam. Since 1948, he has been Professor of Psychology at this university and since 1950, Director of the university's Psychological Laboratory. At present he teaches general and social psychology.

RAGNAR ROMMETVEIT is Associate Professor of Social Psychology at the University of Oslo and Research Director of the Institute for Social Research. He is co-author of *Cross-national Social Research: A Preliminary Frame of Reference*, Oslo, 1951, and author of *Youth and Religious Belief*, Oslo, 1951, and *Social Norms and Roles*, Oslo, 1953.

JOACHIM ISRAEL is on the staff of the Department of Sociology at the University of Stockholm. He holds fil. kand. and fil. lic. degrees from the University of Stockholm. In the past, he has participated in UNESCO's tension project, doing a study of Swedish child rearing practices.

CECILY DE MONCHAUX is Lecturer in the Department of Psychology of University College, London.

THE THREE-DIMENSIONAL BASIS OF EMOTIONAL INTERACTIONS IN SMALL GROUPS. I[1]

F. KRÄUPL TAYLOR

This paper is intended mainly as a report on new methods which have been devised for the purpose of studying quantitatively some of the emotional interactions that develop in small groups. Some results, obtained by the application of these new methods, will also be mentioned.

The sector of group events which will be considered is necessarily a narrow and circumscribed one. There will be no attempt to study processes of verbal, as distinct from emotional, interactions—processes which have been investigated by R. F. Bales (1950) with the help of his "interaction process analysis". Nor will there be any "situation analysis" of the kind reported by Florence B. Powdermaker and Jerome D. Frank (1953) which relates emotional responses among group members to particular group situations of affective significance.

The emotional interactions with which this paper will be concerned, are of a fairly stable kind; they are not the fleeting feelings that are temporarily roused by the changing topics of discussion or by transient group situations. This study will concentrate upon the more or less enduring feelings of friendship or animosity that group members entertain towards their fellow members, and those reciprocal feelings members believe their fellows entertain towards them.

The fact that the group situation in general, and the emotional group field or atmosphere, can exercise an important influence on the emotional reactions among group members—a fact which Lewin, Lippitt, and White (1939) demonstrated experimentally—has not been left out of account in this study. The general situation and atmosphere in most of the groups examined here has been fairly uniform, as I have dealt chiefly with therapeutic groups whose treatment had been conducted on very similar lines,

1. I am greatly indebted to the Board of Governors of the Bethlem Royal and the Maudsley Hospital for a grant from the Research Fund for the purpose of this study.

modelled in part on the principles of group-analytic psychotherapy described by Foulkes (1948). I have given a brief account of my therapeutic approach elsewhere (1952).

CHOICE OF OBSERVER OF GROUP EVENTS

Observed group events vary with the observer collecting the data. In therapeutic groups one has the choice of three types of observer who can be independently employed to record data of emotional interaction: the group therapist, outside observers, and the group members themselves.

Each type of observer has his own merits and demerits; each is qualified to attend particularly to certain aspects of the emotional interrelations in a small group. Ideally, all three types of observer should be enlisted in an investigation because the areas of their observations are often supplementary, and where the areas coincide, an opportunity is offered to check the validity of the data collected. However, for practical reasons it is often advisable to employ only one type of observer.

The majority of papers dealing with emotional phenomena in therapeutic groups have, of course, come from group therapists, who are certainly best qualified by training and experience to observe a wide range of feelings, even those not openly displayed, which have to be inferred from unguarded phrases and undertones. Yet the very range, variety, and equivocation of emotional group events which the therapist has to take into account in conducting a group, makes him an uncertain and often vacillating co-operator in assessing the overt emotions which group members feel towards their companions.

Outside observers have the advantage that they can limit themselves to noting only certain categories of group events. Their observations are often regarded as scientifically superior to the clinical impressions of the group therapist because their records can be compared and tested for reliability. In general it can be said that the reliability of their observations will be the more satisfactory the better their training and the greater their competence; but this reliability will also depend on the conspicuousness of the phenomena they are required to observe. If they were to record vague and variable phenomena—that is, if they were to attend to the kind of data on which therapists often have to base their interpretations—the reliability of their data would be bound to suffer.

The third type of observer—the group member himself—has unique advantages in the task of recording emotions. He does not have to depend on observing other people, nor does he have to make allowance for the fact that people differ in their readiness to display emotions. He can use introspective observation and record even those feelings which he may regard as private and endeavour to conceal from a group companion or an outside observer.

For the purpose of this study I have chosen to rely entirely on observa-

tions by the group members. I felt that in a study aimed at analysing emotional interactions, the introspective assessment of private feelings was indispensable. No outside observer could have given so authoritative an estimate of a feeling as the person who was himself stirred by it.

Yet the choice of group members as observers of emotional interactions may be criticized on the score that introspective assessments are "subjective" and unreliable. The assumption is often made that reliance can only be placed on the recording of data that are "objective" in the sense that they are overt and that they can be observed with equal ease by different persons. It has, for instance, been assumed that the feelings of friendship and animosity which unite or separate individual group members, can be adequately studied by merely noticing "objectively" the number of social contacts spontaneously made by the various group members (provided, of course, that the social field of the examined groups was such that no barriers were placed in the way of such spontaneous contacts).

To obviate a critical disparagement of the introspective observations of group members, it seems advisable to examine more closely the basis of assumptions that the counting of spontaneous social contacts is an adequate indicator of the network of emotional interrelations in a small group. These assumptions seem to derive mainly from two premises:

1. that people tend, in general, to display openly their feelings towards others.
2. that spontaneous social groupings are predominantly motivated by feelings of friendship, and the avoidance of the company of others by feelings of dislike.

There can be little doubt that the first premise is not likely to be correct; on the contrary, it seems—at least in England—that people are, in general, rather reserved in revealing their interpersonal feelings. But even if this premise were true, there is such a wide variation between people in this respect that the feelings of many individuals would be completely misjudged on this basis. Shy and introverted persons would be considered unfriendly and hostile on account of their difficulty in making approaches even towards people they like: on the other hand, sociable and extraverted individuals would be credited with stronger feelings of affection and sympathy than they actually possess.

The second premise is equally suspect. Spontaneous social groupings are not infrequently determined by many factors other than the mere likes and dislikes that grow up among the members of a group. This was clearly shown when H. H. Jennings (1950) investigated the relationship between feelings of friendship and social groupings. She determined, first of all, what she called the "emotional expansiveness" of a person. This concept was operationally defined by the number of people a person listed as desirable companions, choosing from all the members of his group. This number was

compared with the number of people with whom he associated spontaneously in the course of his daily activities ("social-contact range"). Jennings found that, in general, social contacts were initiated with six times as many people as had been listed as desirable companions. There was only a slight correlation of +·22 between "emotional expansiveness" and "social-contact range", though this correlation was statistically significant (P=·01) because of the large population investigated. A comparable observation was made by Polansky, Lippitt, and Redl (1950) who found that people who—unknown to each other—entertained mutual feelings of friendship did not associate with each other more often than with other persons with whom there was no mutuality of private friendship feeling.

Thus we must regard as untenable the two premises on which the assumption rests, that the "objective" recording by outside observers of interpersonal attractions and rejections is superior to the introspective assessment by the subjects concerned of their own interpersonal feelings. In fact, it is difficult to see how a superiority of "objective" observation in this sphere of interpersonal emotions could ever be established. As long as one can rely on the sincerity and cooperation of the subjects who record their own feelings, and as long as only conscious emotions are considered, the validity of introspective data cannot be impugned by evidence derived exclusively from the "objective" observation of manifest behaviour.

The Basic Emotional Variates Considered

The essential data necessary for the evaluation of emotional interactions among group members are of two kinds:

1. The feelings a subject entertains towards each of his group partners. These feelings will range from affection through indifference or ambivalence to hostility. They will be designated by the term "love-hate feelings" to indicate the polar extremes of this emotional variate.
2. The feelings a subject assumes that the various group partners entertain towards him. These assumptions appear to be guesses rather than observations—guesses inspired by the subject's hopes and fears concerning the emotional appeal he has for others, and not firmly based observations of the feelings he has actually aroused in others. This emotional variate will be designated by the term "guessed self-appeal".

Love-hate feelings which are intense and positive are comparable to the sociometric choices elicited by means of Moreno's sociometric test (1951). In a previous attempt at studying the emotional interactions in small groups (1950) I had largely based myself on the sociometric techniques devised and elaborated by Moreno and H. H. Jennings (1945), Lazarsfeld (1945), Bronfenbrenner (1945), and others. It was because the results I had obtained with these sociometric methods were only partially satisfactory that I went on to search for better techniques. Some of the new methods I have been

able to evolve have already been reported in a previous paper (1951), although the terminology used has been drastically altered since then.

I must emphasize, however, that there are essential differences between Moreno's concept of sociometric choice and the concept of love-hate feelings which I am using. Love-hate feelings indicate an emotional variate ranging from strongly positive to strongly negative interpersonal affects. The sociometric choice of an associate, on the other hand, refers to only a limited range of positive feelings. It is true that in some sociometric investigations the subjects had been required to state not only whom they chose, but also whom they rejected; yet the data thus elicited remained fractional data indicating merely narrow ranges of affect near either pole of the emotional variate of love-hate feelings.

It is not surprising, therefore, that sociometrists have experienced difficulties when trying to amalgamate their "choices" and their "rejections" into a single meaningful measure. Lemann and Solomon (1952) admitted this frankly when they stated: "Probably the biggest single gap in the literature on sociometric technique has been the failure to find a measure of social status that utilizes rejections as well as choices". They went on to offer an ingenious solution based on the use of trivariate scatter diagrams.

The restriction of data to sections of a variate introduces unknown elements of biased selection and sampling, and this is bound to make it difficult to evaluate the significance of findings. I shall show later, when discussing the mutuality of interpersonal love-hate feelings, that reliance on fractional sociometric data can give rise to misleading results.

The second emotional variate with which I am concerned, the guessed self-appeal, ranges through the whole gamut of positive and negative feelings that a subject may assume to have aroused in his various group partners. The positive extreme of this emotional variate indicates the subject's conviction that he is loved by a group companion; its negative extreme reveals the subject's belief that a particular group member is his determined enemy who pursues him with intense hatred. In general, guessed self-appeal will, of course, refer to feelings of less extreme intensity.

The assumption that one makes a good or a bad impression on a companion, that one enjoys his affection, or suffers from his ill-will, is certainly one of the prime movers of social behaviour, determining either social ease and confidence, or anxious feelings of inferiority and victimization. Yet these assumptions concerning one's appeal in interpersonal relations seem to have received comparatively little consideration from social psychologists. Tagiuri (1952) has drawn attention to them. He characterized them as social perceptions: "The perception that each person has of the other person's response towards him." He was, however, well aware that he was dealing with "guesses" rather than with the perception of feelings in others. Borgatta (1954) refers to such guesses of self-appeal as "sociometric self-ratings" or "sociometric perceptions".

Both these authors deal, however, with fractional data only. Their subjects were asked to state who, in their opinion, liked them best or least, or enjoyed their company best or least. Their data therefore indicate only narrow sections of the total range of assumptions to which the emotional variate of guessed self-appeal can refer. Thus, their findings are likely to suffer from the same handicaps indicated before with regard to findings based on other fractional sociometric data.

A THREE-DIMENSIONAL FRAME OF REFERENCE FOR EMOTIONAL INTERACTIONS IN SMALL GROUPS

The fundamental aim of my analysis of the emotional interactions in small groups is to reduce their confused and tangled networks to a semblance of order by relating them to a definite framework of social-psychological dimensions.

It is possible to distinguish in a group three kinds of norms or standards which determine the distribution of group variables along three distinct dimensions: (a) a public dimension of variables which have reference to events in the whole group; (b) a dyadic dimension of variables which have reference to events in person-to-person interactions; and (c) an autistic dimension of variables which have reference to events in the intra-psychic sphere.

Public Dimension. Any activity by a group member which is performed in front of the group is a public activity and can be likened to the performance of an actor before an audience. The group members, and particularly the patients in a therapeutic group, are often acutely conscious of this public aspect of their behaviour in the group. Even when not actively participating in group activities, they may yet feel as though on a stage and exposed to the public gaze. Such feelings are excessive, but there is an element of truth in them. The group does, in fact, not only take good notice of the behaviour of its members, it goes further than that and invariably develops a critical and differential attitude towards its various members. This attitude of the group as a whole confers differences of status on the members, and structures the group into a variety of hierarchies. Newcomb (1950) makes a distinction between "ascribed" and "acquired" status (or position, as he prefers to call it). The status that group members achieve through the public aspect of their behaviour is, of course, an acquired status.

In this paper I shall deal with the two public variables which indicate, respectively, the status of popularity and of dominance. As the hierarchical distribution of these variables is a function of the public behaviour of group members, it is plausible to expect that the distribution of public status is common knowledge in the group; or, at least, a phenomenon so conspicuous that it can be readily assessed by "objective" observation. This expectation has, in fact, been largely confirmed by the results of investigations which will be mentioned later.

Dyadic Dimension. The attitude and qualities that characterize the dyadic relations of an individual are independent of the attitudes and qualities that characterize his public relations, and therefore independent of the various forms of public status that he acquires. In his dyadic interactions with different partners a group member may, or may not, display his preferences and aversions, and he may, or may not, be sensitively aware of the differences of feelings which he arouses in his partners. Variables which indicate these dyadic qualities of preference-display and of awareness of one's personal appeal in person-to-person relationships will be dealt with in a future paper. Here my main concern is to give an account of the method of obtaining and measuring dyadic variables, and this can best be done by using as an illustration a dyadic variable that indicates the mutuality of love-hate feelings which different group members achieve.

It must be mentioned that the distribution of dyadic variables in a group depends not only on the particular personalities of the group members, but also on the opportunity which the general group situation provides for making person-to-person contacts, and on the length of time the group has been in existence. Thus, in comparing dyadic variables in different groups, one has to take these two factors—the opportunity for dyadic contact, and the life span of the group—into account. Most of the groups examined in this paper were therapeutic groups, tested at a fairly early stage of their existence. The investigation, however, includes also some student groups which had offered to their members a better and more prolonged opportunity for person-to-person contact than had been the case in therapeutic groups. As expected, the student groups showed a greater degree of mutuality of love-hate feelings.

Autistic Dimension. A person who is required to assess his interpersonal feelings by introspective observation will inevitably be biased by very private and personal tendencies, desires, hopes, fears, and fantasies. A very important role in this respect will be played by his self-estimate of the kind of person he is, wishes to be, or fears himself to be. If, for instance, he has formed an imaginary picture of himself as a friendly and affectionate person, then the total of his scores of love-hate feelings will be greater than that of a person who is afraid to admit friendly ties with other people or who wishes to withdraw emotionally from his group companions. There will be a similar difference in the totals of the scores of guessed self-appeal between a person who believes, desires, or hopes to be popular, lovable and loved in the group, and another person who feels or fears that he is criticized, unloved, and rejected.

The totals of all the scores of love-hate feelings or of guessed self-appeals which a subject has recorded, thus indicate two autistic variables; (a) an autistic bias to profess a particular degree of love-hate feelings in the global relation with the group as a whole, and (b) an autistic bias to profess a belief of arousing a particular degree of love-hate feelings in the group as a whole.

These two forms of global autistic bias will be used as the main illustrations of autistic variables in this paper. As the method of estimating these biases is similar to the method employed in assessing public variables, an account of global bias variables will precede, in the body of the paper, the description of dyadic variables.

Another autistic bias that will receive some attention is a bias that takes account of both the global biases mentioned above. This second-order variable relates to the tendency to assume (or to deny) that one's global emotional attitude towards the whole group is reciprocated by the whole group. This global reciprocity bias is of theoretical interest as it seems to have some of its roots in certain narcissistic fantasies, such as the narcissistic fear of not receiving sufficient libidinal supplies from the group in spite of a liberal cathexis of libido in the group.

Not all autistic bias variables refer to the global relations between ego and total group. An autistic reciprocity bias can also be discerned in the dyadic relations between ego and individual partners; namely, the tendency to assume (or to deny) that the degree of love-hate feelings which is entertained towards different partners is reciprocated by these partners. This dyadic reciprocity bias will not be examined in this paper, but will be reported in a later publication.

A group member who records his love-hate feelings and guessed self-appeals will apply norms derived from all three dimensions. It will therefore be possible to relate the scores he records to variables in these three dimensions. One may, for example, find that a group member A, in recording a particular score of love-hate feelings towards his group companion B, has been affected by such influences as the public popularity status that B enjoys, by his own dyadic tendency to divide the group into friends and foes, and by his autistic fantasy picture of himself as a friendly person.

It is possible to devise methods of measuring variables in the three dimensions, which are more or less independent of the limitations imposed by the characteristics of individual groups of different sizes. The methods applied in performing this task, and some of the results obtained by relating variables thus measured, will form the main body of the paper.

THE GROUP POPULATION INVESTIGATED

The reported results refer to the following groups: 16 therapeutic groups with a total membership of 112 persons (51 men and 61 women), and 4 student groups with a total membership of 39 persons (16 men and 23 women).

The members of the therapeutic groups were neurotic out-patients who attended group sessions once a week. These groups had a membership varying from 6 to 10. They differed with regard to sex composition; some contained only members of one sex, others of both. *Table I* shows the different sex compositions of therapeutic groups.

TABLE I SEX COMPOSITION OF THERAPEUTIC GROUPS

Type of Groups	*Males*	*Females*	*Total*
3 all-male	22		22
5 all-female		32	32
8 male/female	29	29	58
16 therapeutic groups	51	61	112

Two of the student groups were University Extension Classes, and two were student classes of clinical psychologists and psychiatric social workers, respectively. These groups varied in size from 9 to 11. They were all mixed male/female in composition.

The therapeutic groups were examined at a fairly early stage, namely, when the group membership had remained stable for 8 to 12 sessions after the preliminary readjustment of composition which usually follows upon the assembling of a new therapeutic group. When the patients' willingness to co-operate had been obtained, a suitable date for administering the tests was fixed to make sure that, as far as possible, all group members could be present. The tests were given under supervision to prevent any discussion or comparison of replies. The patients were assured that their answers would be confidential. When they had completed the tests, they were not permitted to make any alterations on second thoughts.

The student groups were examined at a much later stage of their existence, but the same precautions were taken to ensure that the test replies were independent and confidential.

THE PUBLIC DIMENSION

It is assumed that the hierarchical distribution of public variables is such a conspicuous phenomenon that the group members—or any other type of observer—cannot help but notice that the group is structured in diverse ways according to such concepts as prestige, popularity, dominance, or leadership.

In order to obtain an estimate of the hierarchies associated with the concepts of popularity and dominance, the group members can be asked to rank their companions with regard to them. From their rank scores it is possible to obtain variables which are then operationally defined, and which will be characterized by capital initials as Public Popularity and Public Dominance to distinguish them from the more vague and general concepts of popularity and dominance.

The instructions concerning the ranking tests of Public Popularity and Public Dominance are set out in *Appendix I.*

If the rank scores of the various group members are arranged horizontally a matrix of a special kind is obtained which may be labelled a

"ranked sociomatrix". An entirely fictitious example of such a matrix is illustrated in *Table II.*

TABLE II *RANKED SOCIOMATRIX*

	A	B	C	D	
A		1	2	3	
B	1·5		1·5	3	
C	1	3		2	
D	2	1	3		
a-scale	4·5	5	6·5	8	Sum= 24
α-scale	75·00	66·67	41·67	16·67	Sum=200

It will be seen that there are no entries in the principal diagonal of this matrix as the group members had been requested to assess the rank positions of their companions, but not their own. Group member B had been undecided whether to accord first rank to A or C. They were therefore given "tied ranks" using the so-called "mid-rank method" (i.e. A and C were given rank numbers corresponding to the average rank score they would have possessed had they been ranked consecutively). This method ensures that the sum of all the rank scores given by any one individual to all his partners (i.e. the sum of all the rank scores in any one row) will be the same whether there are tied rankings present or not. This sum will be $\frac{1}{2}n(n-1)$, when n stands for the number of people in the group.

Had the group members been ideally competent in judging the public hierarchy in question, and had they been entirely impartial and unemotional in their estimates, their rank scores would have exhibited the highest possible concordance. *Table III* shows a fictitious example of such a ranked sociomatrix of maximum concordance.

TABLE III *RANKED SOCIOMATRIX OF MAXIMUM CONCORDANCE*

	A	B	C	D
A		1	2	3
B	1		2	3
C	1	2		3
D	1	2	3	
a-scale	3	5	7	9
α-scale	100·00	66·67	33·33	0·00

Maximum concordance among the group members, who here act as the observers of an "objective" phenomenon, means that they all agree that a particular person occupies a particular status in the hierarchy under consideration. If the observers had not been part of the group, if outside observers

had been used instead, to judge the hierarchical distribution of the public variable, then a maximum concordance would have shown itself in the complete uniformity of all the rank scores in any of the columns of the ranked sociomatrix. In *Table III* this is only the case in the columns of A and of D, who occupy, respectively, the highest and lowest status positions. In the columns of B and C, whose status positions are intermediate, this criterion of maximum concordance is masked by the fact that the group members themselves act as observers, and do not assess their own status so that a blank principal diagonal results in the sociomatrix. In a matrix of this kind another criterion of maximum concordance can, however, be used; namely, that the order of the rank scores is the same in each row. In the sociomatrix of *Table III* the rank scores are in the natural order in each row.

If we add the column-scores in a ranked sociomatrix of maximum concordance we obtain a scale—here called an "a-scale"—which, because it is derived from hypothetical observers of ideal competence and impartiality, reflects accurately the true hierarchical distribution of the public variable assessed.

An analogous a-scale can be obtained by summing the column-scores of the ranked sociomatrix of *Table II*, in which the concordance was not maximum. This a-scale in *Table II* can therefore not be accepted as an entirely accurate measure of the true hierarchical distribution. The question then arises, how closely does it resemble the true hierarchy? Or, to use statistical phraseology, is it the "best available estimate" of the true hierarchy? This question can be answered by statistical standards. If the answer is positive, we are left with the supplementary question whether—by practical standards—the best available estimate is good enough for the purpose at hand.

In order to find a statistical answer, we need, first of all, a measure of the degree of concordance that exists in a ranked sociomatrix. Such a measure, though for matrices in which the principal diagonal was not blank, was devised by Kendall (1948) and called the "coefficient W of concordance". In a previous paper (1951) I have shown how this coefficient can be transformed into a coefficient W′ which indicates the degree of concordance in sociomatrices of the kind with which we are dealing here.

THE COEFFICIENT W′ OF CONCORDANCE

The higher the degree of concordance in a ranked sociomatrix the greater will be the variance of the a-scale of column-totals. The coefficient W′ of concordance is therefore expressed as the ratio between the variance found in a particular a-scale and the highest possible variance an a-scale of the same size can have (i.e. an a-scale derived from a ranked sociomatrix of maximum concordance).

The formula for the coefficient W′ of concordance is

$$W' = \frac{S'}{\frac{1}{12}n^3(n^2-1) - n\sum_{T'} T'}$$

In this formula

n = number of group members;

$S' = \frac{n^2}{(n-2)^2}$ multiplied by the sum of squared deviations from the average in an a-scale;

$T' = \frac{1}{12}\sum_{t}(t^3 - t)$, where t refers to a set of ties in any one row. If, for example, a row contained two sets of ties, of two and three items respectively, T′ would be $\frac{1}{12}(6+24) = 2\frac{1}{2}$. Summing T′ over all rows yields $\sum_{T'} T'$ in the above formula for W′.

The coefficient W′ can range from zero (no concordance) to one (maximum concordance).

The statistical significance of the values of W′ can be satisfactorily assessed by means of the formulae and tables supplied by Kendall (1948), if we replace W and S in his formulae by W′ and S′, and equate m and n. Kendall showed that a statistically significant coefficient of concordance indicated that the a-scale of column-totals is the best available estimate of the true distribution of the phenomenon in question.

In almost all the twenty groups I examined (therapeutic and non-therapeutic), the group members were satisfactorily concordant in their estimates of the Public-Popularity status of their companions, and even more concordant in their estimates of Public-Dominance status. The average value (and standard error) of W′ in estimating Public Popularity was ·57±·03, and in estimating Public Dominance ·72±·04. With one exception (concerning a Public-Popularity estimate), the concordance was always statistically significant and often very highly so. The aggregate significance of the average W′-values is therefore extremely high, and there is no doubt that, in general, the a-scales of Public Popularity and Public Dominance can be regarded as the best available estimates of the true hierarchical distribution of these public phenomena.

But statistical significance is largely a function of the size of the investigated population, and is not a sufficient guide in assessing the practical significance of the values found. We still have to enquire whether these values are good enough for our purpose.

A coefficient of concordance may be regarded as an indicator of "observer reliability". The usual criterion of observer reliability is the mean value of all the rank correlations between the scores of all possible pairs of observers. Kendall (1948) showed that the coefficient of concordance tends to be

slightly higher than this mean value. In our case, which concerns an average W′-value, there is, however, practically no difference.

We may therefore regard the average W′-values of ·57 and ·72 as measures of the average reliability of group members in judging Public Popularity and Public Dominance, respectively. The observer reliability which is thus indicated, is not very impressive. It is not good enough to feel confident about the individual scores in an a-scale. These cannot be accepted as reflecting accurately the real status of individuals, unless the individual happens to have such an extreme and conspicuous status that there is complete, or almost complete, agreement among the group members concerning his rank in the hierarchy. On the whole, however, the average observer reliability, which we found, is only good enough to allow us to accept the general trends of the a-scales of Public Popularity and Public Dominance as practically satisfactory estimates of the true general trends of these hierarchies.

One may make it an arbitrary rule to accept as practically satisfactory only those a-scales of public phenomena which yield a coefficient W′ of concordance of at least ·50. If we do so, five a-scales of Public Popularity and two a-scales of Public Dominance have to be discarded, despite the fact that many of these scales were statistically significant beyond the ·01 level of confidence. In only one group (it happened to be the student class of clinical psychologists) was it necessary to discard both a-scales.

CONDITIONS DETERMINING OBSERVER RELIABILITY

Among the conditions which seem of significance in determining observer reliability, the following may be considered:

1. The competence of the observers in judging the phenomena in question. Group members cannot be regarded as having more than average ability in this respect.

2. Freedom from emotional bias. Group members are likely to be less impartial than outside observers who have no personal stake in group activities.

3. The conspicuousness of the phenomenon judged. The more conspicuous a phenomenon the more likely it is to be assessed with a high degree of concordance even by relatively incompetent and biased observers.

4. The adequacy of the test by which the phenomenon is assessed. The better the instrument of observation the greater is the likelihood of concordance among observers. In our particular case much depends, for instance, on the avoidance of semantic ambiguity and vagueness in the wording and presentation of the tests. It was found—to give just one illustration—that when the group members were asked to rank "leadership qualities", the meaning they attached to this term was so diverse that the resulting concordance of their rank scores was rather unsatisfactory.

The average observer reliability in judging Public Dominance was found

to be superior to that attained in judging Public Popularity. The difference between the two average W′-values is statistically significant beyond the ·01 level ($t=2{\cdot}89$, D.F.$=38$). All the four conditions, listed above, are likely to have played a role in causing this difference. In particular, it seems that Public Dominance is a more conspicuous phenomenon than Public Popularity, and is more uniformly defined in the minds of the group members to whom it appears closely connected with the number of verbal contributions to the group discussions (at least, in the kinds of groups examined here). It is also possible that a large part of the difference in observer reliability is accounted for by the differential effect of emotional bias which is assumed to be more pronounced in the assessment of Public Popularity.

A METHOD OF RENDERING a-SCALES INDEPENDENT OF GROUP SIZE

The scores of an a-scale depend, in part, on the size of the group from which they are derived. *Table III*, for instance, shows that, in a group of four people, the smallest possible a-score is 3, and the highest possible a-score 9. In general, in a group of n people, the maximum range of a-scores extends from $(n-1)$ to $(n-1)^2$.

In order to free a-scores from their dependence on group size and to make them comparable, even if they have been derived from groups with widely different memberships, one can utilize the fact that the maximum range of a-scores is uniquely defined by n [the maximum range is $(n-1)^2-(n-1)=(n-1)(n-2)$]. If we measure a-scores as deviations from the lowest possible value, and express these deviations as percentages of the maximum range, we obtain scores which are of comparable magnitude in groups of any size, and which have a maximum range from 0 to 100. These A-scores, as I shall call them, suffer, however, from a handicap which is due to their derivation from rank scores: the magnitude of a-scores and the magnitude of the status, indicated by them, are inversely related so that the lower the value of the a-score the higher is the status to which it refers, and *vice versa*. This can easily lead to misunderstanding, even if constant circumspection is exercised in presenting and interpreting the meaning of a-scores.

It is, however, not difficult to eliminate this handicap. As the maximum range of A-scores is 100, it is merely necessary to subtract A-scores from 100 in order to obtain what will be called α-scores whose magnitude is directly related to the status they indicate.

The formulae transforming a-scores into α-scores are:

$$\alpha = 100 - A$$

$$A = 100\,\frac{a-(n-1)}{(n-1)(n-2)}$$

This transformation, of course, does not entirely remove all the effects that varying group size can have on a-scores. But the inaccuracies involved

are negligible compared with the great advantage that the α-scores derived from different groups can be put together into a comprehensive α-scale because their magnitudes are comparable.

α-scores can range from 0 to 100. Their average value in groups of any size is always 50.

Assessing One's Own Public Status

We have seen that the group members have a fair knowledge of the public status of others, and therefore achieve a fair degree of concordance when they estimate the Public Popularity of their companions, and an even better degree of concordance when they judge what status of Public Dominance their fellow members occupy.

Let us then consider the related problem of the degree of success group members achieve when assessing their own status in these two public hierarchies.

In order to obtain the necessary data, the group members were asked, after they had ranked their group companions with regard to Public Popularity and Public Dominance, to indicate which position they believed they themselves occupied in these hierarchies. They were enjoined to be candid, and not to place themselves—modestly or hopefully—in an average rank position, unless they thought this was correct. (For details of the instructions given see *Appendix I.*)

Our task was to compare these self-assessments with the corresponding a-scores. For this purpose the device was employed of converting the a-scale into a rank scale (ranging from 1 to n) and of transforming the self-assessments into a rank scale of the same range. The rank correlation coefficient between these two rank scales was then used to indicate how well the members of a group were aware of their own public status.

Only a-scales with a W′-coefficient of at least ·50 were considered. Relevant data were available from 14 scales each of Public Popularity and Public Dominance.

With regard to Public Dominance the average rank correlation coefficient was found to be $+\cdot76\pm\cdot05$, and in the case of Public Popularity it was $+\cdot33\pm\cdot09$. Both these average values are more than three times their standard errors, and therefore highly significant statistically.

The difference between these two average values is also highly significant ($t=4\cdot314$, D.F.$=26$, $P<\cdot001$). If we regard these values as indicators of the average degree of insight group members have into their public status, we come to the conclusion that they are far better aware of their own dominance than of their popularity.

If we compare the average rank correlation of the self-assessments of Public Dominance (i.e. ·76) with the average W′-value (which is comparable to the value of an average rank correlation) when the Public Dominance of

others is assessed (i.e. ·72), we notice the close similarity of the two values. This suggests that group members have approximately the same knowledge of the dominance status of others as of their own.

With regard to Public Popularity, the analogous values are ·33 and ·57. The difference between these two average values is significant at the ·01 level ($t=2{\cdot}831$, D.F.$=32$). It appears therefore that group members are, in general, less well aware of their own popularity than of that of others.

It we again use the arbitrary limit of the value of ·50 to separate practically useful from practically useless scales, then only the self-assessments of Public Dominance can be accepted as having some practical significance, at least as far as average trends are concerned.

To sum up these results: 1. There is evidence that group members have some insight into their own status of dominance and of popularity. 2. From a practical point of view, however, the self-assessments of popularity are too inaccurate, and only the self-assessments of dominance have some validity. 3. Group members have approximately the same insight into the dominance of others as into their own, but they are significantly better aware of the popularity of others than that of themselves.

THE RELATIONSHIP BETWEEN PUBLIC POPULARITY AND PUBLIC DOMINANCE

In 14 groups (therapeutic and non-therapeutic) with a total membership of 108 persons, the a-scales of both Public Popularity and Public Dominance had a W′-coefficient of at least ·50. The respective α-scores of these two public variables were found to correlate +·51. This correlation, though statistically highly significant, demonstrates only a moderate linkage between the variables. It indicates that a person who is very dominant will, more often than not, be also rather popular.

There will, of course, be many exceptions from this moderate tendency which links the two public phenomena, and as these exceptions are often due to the dictates of individual psychopathology, they may be of particular interest to the clinician. Some group members, for instance, are very popular, though they take little part in group activities and are therefore non-dominant. Many of these members have been found to be women with a strong autistic desire for narcissistic supplies who constantly exert themselves to solicit affection even from people with whom they have only a tenuous social relation. There is some hope as we shall see that it may be possible one day to measure this autistic desire with sufficient accuracy to study this problem in quantitative detail.

Of great clinical interest also are some of those group members who are highly dominant and yet unpopular. These members have, as I described in previous papers (1950a, 1951, 1952), a tendency to leave the group prematurely, giving very plausible reasons for their desertion. It also seemed that these dominant, yet unpopular members tended to occupy a particular position in the group which one might call the position of deputy leader,

i.e. they tended to be second in dominance status and to compete for the first place.

The methods by which groups succeed in driving out these particular members are often very subtle, though a group therapist who is aware of this eventuality will have no difficulty in discerning manœuvres of this kind. In groups under autocratic rule these aggressive practices against certain dominant partners or potential leaders are apparently much more obviously displayed. K. Lewin and his associates (1939, 1948) report of two boys whose prestige had been fairly high, but who became, under autocratic rule, the victims of the group's aggressiveness. They both failed to return to the group. Lewin rightly characterises this as a scapegoat phenomenon because the attacked boys were really substitutes for the autocratic leader whose domination could not be challenged.

The desire to find or to engineer scapegoats can also take other forms in therapeutic groups. The chosen victim need not necessarily be one who is dominant but unpopular. An example of a scapegoat of low dominance has been described elsewhere (F. Kräupl Taylor and J. H. Rey, 1953).

The Relationship between a Member's Public-Popularity Status and the Love-Hate Feelings He receives from His Partners

It is now necessary to investigate what relationship, if any, exists between public variables and the network of interpersonal love-hate feelings. The most important problem in this field concerns the variable of Public Popularity. We have seen that the group members have some knowledge of the Public-Popularity status of their companions, and we may therefore frame our question like this: To what extent are the interpersonal love-hate feelings of group members influenced by the status of Public Popularity which their companions occupy?

In order to elicit the love-hate feelings of group members two tests have been applied which have been termed, respectively, the primary ranking test of love-hate feelings and the questionnaire test. The instructions to group members concerning these tests are contained in *Appendix I.*

The *primary ranking test of love-hate feelings* is analogous to the test by which the hierarchy of Public Popularity had been elicited. The group members were requested to rank their subjective preferences for their partners in the group. Their rank scores were arranged in the form of a sociomatrix ("primary rank matrix of love-hate feelings").

As these matrices contain the scores of subjective preferences, we shall expect them to have a smaller degree of concordance than the matrices of Public Popularity which contain estimates of an objective phenomenon. This expectation is indeed borne out. The average W′-value in 20 such primary rank matrices is ·37, compared with the average W′-value of ·57 in the Public-Popularity matrices. The difference is highly significant ($t=3{\cdot}855$, D.F.$=38$, $P={\cdot}0001$). This is proof that the group members

understood and obeyed the different test instructions, and were able to distinguish between the introspective assessment of their own interpersonal feelings towards group partners and their objective estimate of the public status of these partners.

In the *questionnaire test* the group members were not asked to rank their preferences for different group partners, but to consider each partner in turn. They were therefore given (n−1) questionnaire forms and on each form they answered four questions which referred to their love-hate feelings towards a particular partner (the first four questions of the questionnaire shown in *Appendix I*).

The questionnaire replies were scored as follows: a reply indicating positive feelings (or the absence of negative ones) was given the score of +1, a reply indicating negative feelings (or the absence of positive ones) the score −1, and a reply indicating indecision or ambivalence the score zero.

The scores were arranged in matrix form as illustrated in *Table IVa*.

TABLE IV

(*a*) *MATRIX OF LOVE-HATE FEELINGS*

	A	B	C	D	E	*Row-Totals*
A		+2	+4	+4	−1	+ 9
B	+2		+2	−4	−1	− 1
C	+4	+4		+2	0	+10
D	+4	+2	0		−4	+ 2
E	+3	+3	−1	−2		+ 3
Column-Totals	+13	+11	+5	0	−6	+23

(*b*) *SECONDARY RANK MATRIX OF LOVE-HATE FEELINGS*

	A	B	C	D	E
A		3	1·5	1·5	4
B	1·5		1·5	4	3
C	1·5	1·5		3	4
D	1	2	3		4
E	1·5	1·5	3	4	
a-scale	5·5	8	9	12·5	15
α-scale	87·50	66·67	58·33	29·17	8·33

The rows of the matrix of love-hate feelings consist of absolute scores which can range from +4 to −4. The sum of these row-scores varies from person to person (as shown in the right-hand marginal column of *Table IVa*), according to their different autistic bias in introspective self-assessments. Some may consider themselves as friendly disposed towards everybody in the group, and their row-scores would consequently add up to a high

positive total; others may think of themselves as more discriminating in their love-hate feelings, and thus achieve row-totals near zero; and others still may have a negative row-total because they dislike all, or most, of their group companions.

The varying row-totals in a matrix of love-hate feelings are therefore due to the effect of an autistic variable which will be considered in detail later. In the present context this variable is best eliminated as it can have a distorting effect on the α-scores with which we shall here be primarily concerned. This possibility of distortion is due to the fact that there are no scores in the principal diagonal of the matrix so that a person with a tendency to profess high positive scores may inflate the α-scores of his partners whereas his own α-scores may be reduced by the tendency of another member to profess negative scores towards everybody in the group.

The simplest way to eliminate this effect of the autistic variable is to transform the absolute row-scores into secondary rank scores and thus to equalize all the row-totals. This transformation leads to a secondary rank matrix of love-hate feelings (*Table IVb*).

The average W′-coefficient of concordance in these secondary rank matrices of love-hate feelings was found to be ·35; that is, it was of the same size as that found in the primary rank matrices of these feelings.

The column-totals in both the primary and the secondary rank matrices indicate the popularity which group members enjoy in direct interpersonal relations. We may therefore speak of a variable of Interpersonal Popularity.

We have two estimates of this variable in the α-scales derived, respectively, from primary and from secondary rank matrices. If we assume that the tests by which these two estimates were obtained are of approximately equal power as measuring instruments of Interpersonal Popularity, we shall expect a high correlation between the two α-scales. The correlation is, in fact, +·88.

This indicates a fairly high test reliability. By adding the two α-scales we could further improve it to the very satisfactory value of +·94 (attributing equal weight to the primary ranking test and the questionnaire test, and applying the Spearman-Brown prophecy formula). However, instead of merely adding the two α-scales, I have preferred the more exact way of adding, in each group, the primary and secondary rank matrices. In the resulting matrix the row-scores were re-converted into rank scores. This final matrix may be termed a combined rank matrix of love-hate feelings. The average W′-value in these matrices was ·37. The α-scale derived from them will be used as the best available estimate of the distribution of Interpersonal Popularity.

By correlating Interpersonal Popularity and Public Popularity (in those groups in which a satisfactory degree of concordance in judging Public Popularity existed) we can obtain an answer to our original question, namely, the extent to which the interpersonal feelings of group members are influenced by the Public-Popularity status of their partners.

The correlation coefficient was found to be +·81 (N=115). This indicates that group members are influenced in their feelings for group companions by their conscious or unconscious awareness of the popularity which these companions publicly enjoy in the group, and that this influence is so marked that it almost completely obscures the fact that genuinely personal feelings, unaffected by the partners' public status, do exist in many dyadic relationships. Indeed these personal dyadic feelings must be regarded as mainly responsible for the rather low average W′-value of ·37 in the rank matrices of love-hate feelings. However, these genuinely personal feelings that members record towards a particular partner are obviously so diverse in quality and intensity that they cancel themselves out when that partner's Interpersonal Popularity is calculated by summing all the ranked love-hate scores he had received.

It seems that the influence which Public Popularity exerts on the network of interpersonal love-hate feelings is not the same in all-male and in all-female therapeutic groups. In three all-male groups with a total membership of 22 the correlation coefficient between Public and Interpersonal Popularity was only +·62; whereas in three all-female groups with a total membership of 19 it was as high as +·93. The difference between these coefficients is significant at the ·01 level of confidence. In fact, in each of the three all-female groups the coefficient was over +·90 (·97, ·92, ·93); in none of the three all-male groups did it reach +·80 (·79, ·56, ·52). In nine male/female groups (therapeutic and non-therapeutic) the correlation varied from +·62 to +·98, and in all these groups with a total membership of 74 it was +·82.

This finding suggests that women, when in groups composed of members of their own sex only, model their interpersonal love-hate feelings to a much greater extent on the Public-Popularity status of their partners than is the case with men in all-male groups.

The fairly close parallelism between the distributions of Interpersonal and of Public Popularity in male/female groups brings with it a practical advantage. In these groups it is sometimes of interest to discriminate between the public popularity which group members enjoy with their own and with the opposite sex. These two types of popularity we shall distinguish by the adjectives "homo-social" and "hetero-social"—terms which were introduced by Flugel (1927) to indicate the social aspects of homo- and hetero-sexual relationships.

We cannot, however, obtain measures of homo-social and hetero-social popularity from matrices of Public Popularity as these deal exclusively with the status achieved in the whole group audience of both men and women, But some measure of these sex-linked popularities can be derived from combined rank matrices of love-hate feelings as it is possible to subdivide them into sections containing only homo-social or hetero-social relationships. Such a sub-division is schematically illustrated in *Table V*.

A man's a-score of homo-social popularity is obtained by adding all the

TABLE V *SCHEMATIC ILLUSTRATION OF THE HOMO- AND HETERO-SOCIAL SECTIONS OF A RANK MATRIX*

	Male Homo-social			*Female Hetero-social*			
	a-scale						
	m	*m*	*m*	*f*	*f*	*f*	*f*
m		mm	mm	mf	mf	mf	mf
m	mm		mm	mf	mf	mf	mf
m	mm	mm		mf	mf	mf	mf
	Male Hetero-social			*Female Homo-social*			
	a-scale						
f	fm	fm	fm		ff	ff	ff
f	fm	fm	fm	ff		ff	ff
f	fm	fm	fm	ff	ff		ff
f	fm	fm	fm	ff	ff	ff	

Code: mm = feelings of men towards men
ff = ,, ,, women towards women
mf = ,, ,, men towards women
fm = ,, ,, women towards men

mm-entries in his column, and his a-score of hetero-social popularity by adding all the fm-entries. A woman's a-scores of homo- and hetero-social popularity are obtained by adding, respectively, the ff-entries and the mf-entries in her column.

The conversion of these sex-linked a-scores into α-scores is described in *Appendix II.*

These homo- and hetero-social α-scales have been accepted as the best available estimates of the public variable of same-sex and other-sex popularity, provided that two—rather arbitrary—conditions are fulfilled: 1. that only the scores of those groups should be used which have a W′-coefficient of Public Popularity of at least ·50, and 2. that there should be at least three members in the male or female section of the group.

These conditions were present in 9 of my male/female groups (therapeutic and non-therapeutic) with a total membership of 34 men and 40 women. The average α-score of homo-social popularity in all the 74 group members was $55{\cdot}07 \pm 2{\cdot}83$, and the average α-score of hetero-social popularity $45{\cdot}97 \pm 2{\cdot}70$. The difference between these averages is significant at the ·03 level ($t = 2{\cdot}298$, D.F.$= 146$). If the differences between the corresponding homo- and hetero-social α-scores of the individual members are considered, and a single-sample t-test applied, a significance level of almost ·001 is attained ($t = 3{\cdot}024$, D.F.$= 73$).

We may conclude therefore that a slight, but distinct tendency exists in small groups which renders members, whether male or female, more popular with their own than with the opposite sex.

The correlation between homo- and hetero-social popularity was found

to be +·41 (N=74). This is highly significant statistically, but does not indicate a close correspondence between the degrees of popularity which members achieve with the male and female sections of their groups.

GUESSED SELF-APPEALS AND THE VARIABLE OF ATTRIBUTED INTERPERSONAL FRIENDLINESS

So far I have only considered love-hate feelings. A subject's guessed self-appeals (i.e. his estimates of the feelings or preferences he believes himself to have aroused in his partners) can be elicited by means of a primary ranking test and a questionnaire test (see *Appendix I*).

The primary ranking test yields primary rank matrices of guessed self-appeals. The columns of these matrices indicate the degree of interpersonal friendliness which is attributed to individual members by the group members generally. The a-scores of column-totals in these matrices thus correspond to the distribution of a variable which one may call the variable of Attributed Interpersonal Friendliness. These a-scores can be transformed into α-scores to make their values independent of group size.

The questionnaire test (the last two questions of the questionnaire shown in *Appendix I*) yields a matrix of guessed self-appeals (comparable with the matrix of *Table IVa*) which can be transformed into a secondary rank matrix of guessed self-appeals (comparable with the matrix of *Table IVb*) by converting the absolute row-scores into rank scores.

The α-scales of Attributed Interpersonal Friendliness derived, respectively, from primary and secondary rank matrices of guessed self-appeals, have a correlation coefficient of +·77, indicating a test reliability of ·87 in a combined scale. By adding the primary and secondary rank matrices, and re-ranking the resulting row-scores, combined rank matrices of guessed self-appeals are obtained. The α-scale derived from them will be used as the best available estimate of the distribution of Attributed Interpersonal Friendliness.

One might expect that this variable of Attributed Interpersonal Friendliness is related to a public phenomenon, namely, the degree of general friendliness publicly exhibited by the various group members. There are, however, some indications that such an expectation is likely to be disappointed and that the relationship between Attributed Interpersonal Friendliness and publicly exhibited friendliness will not be found to be as close as the relationship between Interpersonal and Public Popularity.

One such indication is that the average coefficient W′ of concordance is only ·27 in the combined rank matrices of guessed self-appeals. This is significantly lower than the corresponding W′ of ·37 in the combined rank matrices of love-hate feelings (t=2·46, D.F.=38, P=·02). The fact that the concordance of guessed self-appeals is so low suggests that group members base their guesses chiefly on considerations of their dyadic relations and fail to take adequate account of any public phenomenon which—as in the case of love-hate feelings—might have introduced some semblance of agreement

into the divergencies of their dyadic scores. Whatever general friendliness a partner exhibited to the group as a whole, seems to have impressed his companions less than the emotions he displayed—or was felt to have displayed—in his private dyadic dealings. There is also some evidence (which will be reported in the section on the Autistic Dimension) that the divergencies of guessed self-appeals derive in part from autistic sources; namely, from noticeable differences of bias among group members in estimating their global appeal to the group audience.

The correlation coefficient of Attributed Interpersonal Friendliness with Public Popularity is +·62 (N=115), and with Interpersonal Popularity +·63 (N=151). As we do not know to what extent Attributed Interpersonal Friendliness represents a public phenomenon of overtly expressed friendliness, we cannot interpret the first of the above coefficients as indicating either that people who are overtly friendly gain public popularity, or that people who are publicly popular are regarded as friendly towards everybody. But the second of the above coefficients refers to interpersonal—or more exactly, dyadic—data, and this suggests the possibility that the marked positive correlation is due to the working of an autistic influence, namely, of a bias to assume reciprocity of love-hate feelings in dyadic relations. Such a bias, which would tend to equalize the scores of a person's love-hate feelings and guessed self-appeals, undoubtedly exists, but its detailed consideration will be reserved for a later publication.

Homo- and hetero-social α-scales of Attributed Interpersonal Friendliness can be obtained by a procedure similar to that employed in the case of Interpersonal Popularity. The only limiting condition that was imposed, was that there should be at least three members in the male or female section of a male/female group. This excluded only one of our twelve male/female groups. The average homo-social α-score for the 41 men and 50 women in these groups was 57·25±2·29; the average hetero-social α-score was 44·41±1·94. The difference between these averages is highly significant ($t=4{\cdot}282$, D.F.=180, $P<{\cdot}0001$).

There is, therefore, a slight but definite tendency among both the male and female members of male/female groups to attribute a greater degree of friendliness to members of their own sex. The correlation coefficients between homo- and hetero-social Attributed Interpersonal Friendliness was +·23 which, though significant at the ·03 level, does not suggest that there is much agreement among the male and female sections of a group concerning the friendliness they attribute to a particular group.

SUMMARY

This paper deals essentially with new methods of analysing the emotions among members of small groups. Some results obtained in 20 groups with a total membership of 151 have been reported.

The basic emotional variates considered were: (a) the "love-hate feelings" which a subject entertains towards each of his group partners, and (b) his "guessed self-appeals", i.e. the love-hate feelings which a subject assumes the various group partners entertain towards him.

The intensity and the quality of the feelings which a subject entertains or assumes to have aroused in others, was estimated through introspective observation by the group members themselves.

Love-hate feelings and guessed self-appeals were analysed by relating them to a three-dimensional framework in accordance with three distinct norms or standards which exist in a group setting; (a) *A Public Dimension*—containing variables, which derive from the overt behaviour of group members in front of the group audience, and which structure the group along various hierarchies of acquired status; (b) *A Dyadic Dimension*—containing variables which derive from the behaviour of group members in their dyadic interactions with individual partners; and (c) *An Autistic Dimension*—containing variables derived from autistic emotional attitudes towards the group and its members, and from autistic expectations and fantasies about the attitude of the group and its members towards oneself.

Tests of estimating variables in each of the three dimensions are described, and methods of measuring them are reported which exclude, as far as possible, the effect of different group sizes and of other distorting influences.

Two public variables were examined: Public Popularity and Public Dominance. The validity and reliability of the estimates of these variables was assessed by means of a coefficient W′ of concordance. Data which did not reach a satisfactory level of reliability were excluded.

Group members were, in general, fairly successful judges of the public hierarchy of dominance in their group, and also of their own position in that hierarchy. They were significantly less successful in assessing the public popularity of their partners, though in most groups their estimates were of adequate reliability. They were least aware of their own status of public popularity, but they were not completely in the dark about it.

The correlation between Public Popularity and Public Dominance was +·51. The possible psychodynamic significance of discrepant statuses in these public hierarchies has been considered.

The love-hate feelings of group members are influenced by the Public-Popularity status of their partners to such an extent that the marked diversity of genuinely personal feelings in many dyadic relationships tends to be obscured when the scores of Interpersonal Popularity are calculated. As a result, publicly popular people also tend to appear interpersonally popular. This was significantly more true of women than of men, at least in one-sex groups.

Guessed self-appeals are apparently less affected by the public status of partners.

REFERENCES

BALES, R. F. *Interaction Process Analysis.* Cambridge, Mass.: Addison-Wesley Press. 1950.

BORGATTA, E. F. "Analysis of Social Interaction and Sociometric Perception." *Sociometry*, 1954, **17**, 7–32.

BRONFENBRENNER, U. "The Measurement of Sociometric Status, Structure and Development." *Sociometry Monographs*, No. 6. New York: Beacon House, 1945.

FESTINGER, L. "The Analysis of Sociograms Using Matrix Algebra." *Human Relations*, 1949, **2**, 153–158.

FLUGEL, J. L. "Sexual and Social Sentiments." *Brit. J. Med. Psychol.*, 1927, **7**, 139–176.

FOULKES, S. H. *Introduction to Group-Analytic Psychotherapy*, London: William Heinemann Medical Books. 1948.

JENNINGS, H. H. *Leadership and Isolation: A Study of Personality in Inter-Personal Relations.* New York, London, Toronto: Longmans, Green & Co. 1950.

KENDALL, M. G. *Rank Correlation Methods.* London: Charles Griffin. 1948.

LAZARSFELD, P. Quoted by MORENO, J. L., and JENNINGS, H. H., 1945 (q.v.).

LEMANN, Th. B., and SOLOMON, R. L. "Group Characteristics as Revealed in Sociometric Patterns and Personality Ratings." *Sociometry*, 1952, **15**, 7–90.

LEWIN, K. *Resolving Social Conflicts.* New York: Harper & Brothers, 1948.

LEWIN, K., LIPPITT, R., and WHITE, R. K. "Patterns of Aggressive Behaviour in Experimentally Created Social Climates." *J. Soc. Psychol.*, 1939, **10**, 271–299.

MAUCORPS, P. H. "A Sociometric Inquiry in the French Army." *Sociometry*, 1949, **12**, 46–80.

MORENO, J. L. *Sociometry, Experimental Method and the Science of Society.* New York: Beacon House, 1951.

Idem. " Sociometry and the Cultural Order." *Sociometry*, 1943, **6**, 299–344.

MORENO, J. L., and JENNINGS, H. H. "Sociometric Measurement and Social Configurations." *Sociometry Monographs*, No. 3. New York: Beacon House, 1945.

NEWCOMB, TH. M. *Social Psychology.* New York: Dryden Press, 1950, London: Tavistock Publications, 1951.

POLANSKY, N., LIPPITT, R., and REDL, F. "The Use of Near-Sociometric Data in Research on Group Treatment Processes." *Sociometry*, 1950, **13**, 39–62.

POWDERMAKER, FLORENCE B., and FRANK, JEROME D. *Group Psychotherapy: Studies in Methodology of Research and Therapy.* Cambridge, Mass.: Harvard University Press, 1953.

TAGIURI, R. "Relational Analysis: An Extension of Sociometric Method with Emphasis Upon Social Perception." *Sociometry*, 1952, **15**, 91–104.

TAGIURI, R., BLAKE, R. R., and BRUNER, J. S. "Some Determinants of the Perception of Positive and Negative Feelings in Others." *J. Abn. Soc. Psychol.* 1953, **48**, 585–592.

TAYLOR, F. KRÄUPL. "The Pattern of Friendliness and Dominance in a Therapeutic Group." *J. Ment. Sc.*, 1950, **96**, 407–425.

Idem. "Quantitative Evaluation of Psycho-Social Phenomena in Small Groups." *J. Ment. Sc.*, 1951, **97**, 690–717.

Idem. "On Some Principles of Group Therapy." *Brit. J. Med. Psychol.*, 1952, **25**, 128–134.

TAYLOR, F. KRÄUPL, and REY, J. H. "The Scapegoat Motif in Society and Its Manifestations in a Therapeutic Group." *Internat. J. Psycho-Analysis*, 1953, **34**, 253–264.

APPENDIX I

TESTS ADMINISTERED

1. Questionnaire Test

You are asked to fill in this questionnaire. The questions are about your thoughts and feelings regarding each of the other group members.

Write the name of the group member you are considering on the top of the sheet. Each question can be answered "Yes" or "No". If your answer is "Yes", put a ring round Yes, like this (Yes). If your answer is "No", put a ring round No, like this (No). If you are undecided, put a ring both round Yes and around No, like this (Yes) (No).

Make sure that you answer every single question.

Name of group member considered:.....

1. Would you choose him (or her) as a friend; that is to say, would you like to meet him (or her) regularly and not just at group sessions?	Yes	No
2. Even if you do not choose him (or her) as a friend would you say that you like him (or her) as a person?	Yes	No
3. Do you dislike him (or her) as a person?	Yes	No
4. Do you think that he (or she) is a pleasant person?	Yes	No
5. Do you think that he (or she) likes you?	Yes	No
6. Do you think that he (or she) dislikes you?	Yes	No

2. Primary Ranking Test of Love-Hate Feelings

You are now asked to put down the names of the other group members in the order in which you like them. Thus you put down first the name of the group member whom you like most. Underneath put the name of the group member whom you like second-best, and so on, until you have considered all the other members in the group.

You may feel uncertain about the place some group members should occupy in the list. For example, you might wonder whether to put Smith or Jones in the 4th place. In that case you should put one of them into 4th and one into 5th place and add an asterisk * after both their names. It would look like this:

4. Smith*
5. Jones*

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

3. Primary Ranking Test of Guessed Self-Appeal

Now put down the names of the other group members in the order in which you think they like you. Thus you put down first the name of the group member who, you think, likes you best. Underneath put the name of the group member who, you think, likes you second-best, and so on, until you have considered all the members in the group.

If you are undecided use asterisks as before.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

4. Ranking Test of Public Popularity

Now I would like to have your estimate of the general popularity of each member in the group. In doing this you should not allow your own feelings towards a particular group member to influence your judgement of his popularity. You might, for instance, dislike a particular person but realize that he is popular with the others. It is a person's popularity in the group generally which I would like you to estimate.

Please put down the names of the group members in order, beginning with the person who appears to you to be the most popular person in the group.

If undecided use asterisks as before.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

5. Self-Assessment of Public Popularity Status

Would you now kindly estimate your own popularity in the group. State what place you think your name will occupy generally in the lists of popularity filled in by the other members. Thus, if you think you are the most popular member put down the figure 1; if you think you are second in popularity put down the figure 2, and so on.

Please give your candid estimate. Do not just put down a figure which corresponds to the middle of the group unless you think that figure is correct. Look up the list of popularity which you have just drawn up and ask yourself where your own place would be.

Estimated own place of popularity: .....

6. Ranking Test of Public Dominance

Now put down the names of the other group members according to the degree to which they have, in your opinion, dominated and influenced the group discussions. Do not allow yourself to be influenced in your estimate by feelings that a particular member's influence on the group had been more harmful than helpful. If you feel that a particular member had the greatest influence on group discussions put his name down first, no matter whether you generally disagreed with him or found him unpleasant.

The name of the person who was second in his influence on the group you put down next, and so on, until you have considered all the members in the group.

If you are undecided use asterisks as before.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

7. Self-Assessment of Public-Dominance Status

Would you now kindly estimate your own influence on the group. If you think you are the most dominant and influential group member, put down the figure 1; if you think you were second in group influence put down the figure 2, etc. Please give your candid opinion and do not just put down a middle figure.

Estimated own place of dominance: .....

THANK YOU.

APPENDIX II

Converting Homo- and Hetero-Social a-Scores into α-Scores

The derivation of homo- and hetero-social a-scales is illustrated in *Table V* on the model of a rank matrix. (From a bias matrix* homo- and hetero-social a-scales are obtained in an analogous manner by adding row-scores, instead of column-scores, in the relevant matrix section.)

Let us use the following subscripts to indicate the various kinds of homo- and hetero-social scores that can be derived from male/female groups:

mm—to indicate male homo-social scores
ff —„ „ female „ „ „
mf —„ „ male hetero-social „
fm —„ „ female „ „ „

Thus a_{mm}- and a_{ff}-scores are obtained from a rank matrix by adding column-scores in the mm- and ff-sections of the matrix, respectively (see

Table V); they are obtained from a bias matrix* by adding the row-scores in the mm- and ff-section, respectively.

a_{mf}-scores are obtained from a rank matrix by adding the column-scores in the fm-section of the matrix; they are obtained from a bias matrix* by adding the row-scores in the mf-section.

a_{fm}-scores are obtained from a rank matrix by adding the column-scores in the mf-section of the matrix; they are obtained from a bias matrix* by adding the row-scores in the fm-section.

In a group of n members, containing m men and f women, the following formulae are applicable:

Converting male homo-social a-scores into α-scores:

$$\alpha_{mm} = 100 - A_{mm}$$

$$A_{mm} = 100\,\frac{a_{mm} - (m-1)}{(m-1)(n-2)}$$

Converting female homo-social a-scores into α-scores:

$$\alpha_{ff} = 100 - A_{ff}$$

$$A_{ff} = 100\,\frac{a_{ff} - (f-1)}{(f-1)(n-2)}$$

Converting male hetero-social a-scores into α-scores:

$$\alpha_{mf} = 100 - A_{mf}$$

$$A_{mf} = 100\,\frac{a_{mf} - f}{f(n-2)}$$

Converting female hetero-social a-scores into α-scores:

$$\alpha_{fm} = 100 - A_{fm}$$

$$A_{fm} = 100\,\frac{a_{fm} - m}{m(n-2)}$$

There is a check on the calculation of all the A-scores, because

$$A_{mm} + A_{ff} + A_{mf} + A_{fm}$$

$$= \frac{100}{f(m-1)(n-2)}\left[f\sum_{m=1}^{m} a_{mm} + (m-1)\sum_{m=1}^{m} a_{mf} - 2mf(m-1)\right]$$

$$+ \frac{100}{m(f-1)(n-2)}\left[m\sum_{f=1}^{f} a_{ff} + (f-1)\sum_{f=1}^{f} a_{fm} - 2mf(f-1)\right].$$

* See the section on the Autistic Dimension.

BIOGRAPHICAL NOTE

DR. F. KRÄUPL TAYLOR took his medical degree in Prague in 1929. He came to Britain in 1939 where he had the opportunity to engage in whole-time private studies of psychology, mathematics, and statistics for two years. In 1942 he took the Diploma in Psychological Medicine at the University of London. At present he is Consultant Psychiatrist at the Bethlem Royal and Maudsley Hospital, and a Clinical Teacher on the staff of London University, lecturing at the Institute of Psychiatry and at the London School of Economics and Political Science. His researches have been mainly concerned with the structure and dynamics of small groups and with the techniques of group psychotherapy.

THE ACCEPTANCE OF NEW BEHAVIOR IN CHILD-REARING*

ORVILLE G. BRIM, JR.

This paper reports a study which was concerned with change in a child-rearing practice on the part of a small group of mothers. The research was not focused upon the specific relationship between some single experimental variable and the change which occurred. Instead, its purpose was to contrast the mothers who changed with those who did not, and to try to bring out of this comparison a full picture of the cluster of variables which were relevant to the process of change.

The innovation studied is part of a current cultural trend toward permissiveness in child-rearing, and this fact provides the study with a broader setting, and makes the results of interest to the field of culture change. The underlying premise is that whatever the factors influential in these individual cases, they are similar to those affecting other mothers in comparable situations, and thus represent in microcosm the factors influencing the adoption of the innovation into the culture at large.

THE RESEARCH PROBLEM

The innovation selected for study is *lenient or permissive behavior on the part of mothers in handling food refusals of children.* Stated more briefly it is the permissive treatment of "feeding problems".

The current widely held view (1, 4, 8, 11, 15) is that feeding problems are caused by oversolicitous or "forced feeding" behavior by a child's parents. New ideas on child care and new information about feeding of children have led to the emergence of a body of advice concerning the handling of food refusals, the essential content of which is that the correct parental behavior when a child refuses to eat is to leave the child alone.

* Part of a dissertation presented to the Faculty of the Graduate School of Yale University in candidacy for the Degree of Doctor of Philosophy. I wish to acknowledge the assistance received in this study from the Social Science Research Council and from the Committee on Child Health of the American Public Health Association. The former provided a research training fellowship during the course of the study, and the latter provided financial aid to meet certain expenses.

The parent should serve the child nourishing and appetizing food, and then allow him to eat as much or as little as he wishes. It is believed that this new advice, if put into practice, not only prevents the development of feeding problems, but has therapeutic value in established cases.

It was necessary to find as subjects for the study mothers who were faced with a feeding problem in their own children, and who had never treated the problem permissively. Permissive behavior as a feeding method could then be introduced to them, and the subjects' acceptance or rejection of the behavior followed through.

Fifty-seven mothers were selected from the records of those attending the Child Health Stations of the New York City Department of Health,[1] according to the following criteria: 1. she was not working, and was taking full care of her child; 2. she was not divorced or separated from her husband; 3. she was either Italian, Negro, or Jew, married to a husband of the same cultural background; 4. she had an eldest child, in good health, between the ages of 1½ and 3 years; 5. she considered this child to present a feeding problem, and was forcing him to eat when he refused his food.

These fifty-seven subjects were interviewed in the Child Health Stations at one of their regular appointments, shortly after their selection. Fifty of these were re-interviewed three to four months later, in the spring of 1951.

During the initial interview information was collected on factors considered relevant to the problem, and current feeding practices were discussed. Only after this, at the close of the interview, was it suggested to each mother that she adopt permissive behavior as a method of handling food refusals. This suggestion, presented in a standardized manner, involved the following: 1. that doctor's believed that such behavior might be helpful in getting the subject's child to eat better; 2. that if a child is allowed to eat what he wishes of good foods presented to him, he will, in general, select for himself an adequate diet over a period of time; and 3. that it often takes two weeks of continued employment of this type of behavior before any increase in eating occurs. At the close of the interview each mother was given a pamphlet describing in detail the permissive feeding practice.

The new behavior pattern had been an effective remedy in many previous cases of feeding problems, and most certainly was superior to the means which the mothers were using at the time it was suggested to them. However, of the 50 mothers who were re-interviewed, 26 did not even try the novel behavior. Of the 24 who did try, 16 rejected it after a few trials. Only 8 of the 50 adopted the behavior as their customary response to their children's food refusals.

In the analysis which follows we shall concentrate on two major topics. The first is the mothers' customary behavior at the time of the first interview.

1. These Stations, which are "well-baby" clinics similar to many found in the United States, serve more than one hundred thousand children each year.

Knowledge of the foundations of the existing behavior with which the novel pattern competed provides the necessary background for an understanding of change.

The second topic is the process of adoption of the permissive pattern by the eight women who made the change. Our purpose here will be to identify and consider the effect of those factors responsible for the change in behavior.

THE THEORETICAL SETTING

The subsequent analysis is carried out in terms of simple expectancy theory, based in the main on the writings of French (3), Hilliard (5), Hull, (6, 7), Parsons (9), Rapaport (10), Tolman and Brunswik (13), and Troland (14). The central premise is that individuals under specified conditions are motivated toward some goal, and select, for performance, behavior which they expect will best attain that goal.

The interpretation of behavior by such a scheme has, as its minimum prerequisite, the designation of the motives present in the conditions, their strength at that time, and the individual's beliefs or expectations about the consequences of various alternatives available in regard to the achievement of the goal or goals.

The motivational basis of any behavior pattern is complex. Though one may perhaps distinguish some specific or dominant motive, it is necessary to consider so far as possible the complex of motives from which the behavior springs. In this particular research the concerns shared by the mothers, in addition to the desire that their children should eat more, were conceptualized as 1. a concern over maintaining the children's emotional contentment or equanimity; 2. a concern over having the children develop a character or personality acceptable to the groups in which they must live; and 3. a concern based on the ubiquitous need for social approval, namely, a desire that their husbands and neighbors should approve of their methods of child-feeding. On only one of these, the desire for the children's contentment, was any attempt made to measure differences in strength of the motive.

The subjects' expectations about the consequences of feeding methods as they relate to these four concerns may be assessed either directly, e.g., simply by asking them, or indirectly by inference. We have made use of both techniques, relying upon one or the other as the situation and the materials dictated.

The data obtained fall far short of those necessary for a rounded understanding of behavior where such a theoretical scheme is employed. Nevertheless they give us a partial picture both of the mothers' actions and of the way in which such an interpretation can be applied to complex social behavior.

In our terms, the mothers' customary feeding behavior at the time of the first interview must be understood to be the result of their efforts to utilize

a method of child-feeding which they expected would be most rewarding to their nexus of interests; that is, a method which they thought would get the child to eat, would keep him well and happy, and would at the same time be socially approved.

In considering the change by eight of the mothers from this habitual behavior to permissive practice, we assume that no change in motivation occurred. Therefore we should concentrate on the factors which altered the subjects' expectations about the results of the various alternative feeding methods available to them, so that permissiveness was tried, continued to be employed, and eventually became substituted for the previous customary behavior.

THE CUSTOMARY BEHAVIOR

During the first interview with the mothers, information was collected on their performance of thirteen different methods of feeding children, ranging from permissiveness at one extreme to so-called forced feeding, i.e., forcibly opening the child's mouth and putting the food in, at the other. A Guttman-type scale analysis was made of the subject's frequency of performance of these thirteen patterns, and three major types of behavior were constructed.[2] These types are referred to as the *aggressive*, *intermediate*, and *persuasive* modes of feeding behavior, and the subjects who customarily used these modes of behavior are termed aggressive, intermediate, and persuasive-type mothers.

The aggressive mothers either beat their children for not eating, or both beat them and forcibly opened their mouths to put in the food. This behavior occurred at least once a week, and its customary performance distinguished these mothers from the others. These same women, however, also yelled at or threatened their children at least weekly and, moreover, used toys or stories or games at meal-time weekly or more often.

The intermediate-type mothers threatened or yelled at their children for not eating, but did not beat them or force their mouths open except perhaps very rarely. In addition to yelling at their children at least weekly, they too employed games, toys, and so forth.

Finally, some of the women used only the persuasive behavior of telling stories, playing games, etc., although this happened at least weekly. Mothers who used only this mode of behavior, and did not threaten or hit the child, are termed the persuasive-type mothers.

In terms of the scale, the behavior may be described as follows: all of the mothers used persuasive behavior at least weekly. Some mothers went farther and yelled at or threatened their child once a week or more often. Of these,

2. Space prohibits presentation of the full supporting data for the scale analysis. The procedure followed is given in S. Stouffer *et al. Measurement and Prediction* (12), pp. 3–45. The reproducibility of the scale was $1-4/(5 \times 57)$, or 98·6.

some went still farther and either beat their child or both beat him and forcibly opened his mouth at least weekly. The distribution of the fifty-seven mothers among these three types is as follows:

Aggressive: 28 cases
Intermediate: 12 cases
Persuasive: 17 cases

The mothers had made little or no change in their customary responses to the meal-time situation since the onset of the feeding problem in their children. They had committed themselves to their current mode of behavior almost at the very start; experimentation with other patterns had been capricious, uncommon, and quite brief. Persuasive behavior constituted the base line of response for the subjects; i.e., eschewing permissive behavior, all of the subjects began at the next level which is the first step toward coercion. The analysis which follows is directed specifically to the difference between those mothers who remained at the persuasive level and those who went on to aggressive behavior.

The differences between the two extreme types of persuasive and aggressive mothers may be quickly summed up in the following manner. The aggressive mothers, who were more severe in their behavior, were more dominant in their child-rearing in general, and at the same time were less constrained by beliefs that such severe behavior would decrease the amount eaten, or harm the child's personality, or bring censure from their husbands. They were, in addition, less perturbed by the emotional upsets that such behavior caused in their children. These differences between the aggressive and persuasive mothers led each group to its customary behavior and served to hold it in line.

The data which follow and support these statements were collected during the interview, and constitute verbal responses to fairly direct questions by the researcher. The context of the interview necessitated such a direct approach rather than the use of more subtle tools of inquiry, and it is highly important in this regard to know that all questions to the mothers concerning their child-rearing beliefs and practices were put prior to the questions about their specific feeding practices, which they had no reason to expect would follow.

The first factor to be considered is the mothers' expectations about the consequences of aggressive feeding behavior. The stated beliefs of the aggressive-type mothers about the consequences for their children's eating showed that they expected better results from such behavior than the persuasive mothers did. More than half of the former held that aggressive behavior either increased or had no detrimental effect on the amount their children ate. This is shown in *Table I.*

At the same time there is evidence that expectations less easily verbalized were also favorable to aggressive behavior. The failure of all of the subjects

TABLE I *EXPECTED EFFECT OF AGGRESSIVE BEHAVIOR ON EATING, BY TYPE OF MOTHERS**

Type	*Increase or No Effect*	*Decrease*
Aggressive	15	13
Persuasive	2	15
		P(E) ·01

* In the statistical analysis of the data in this paper fourfold tables commonly are used. The P(E) designation represents Fisher's exact test. The exact calculation has been carried out wherever the actual or the expected frequency in any cell in the table is less than five. (See Fisher, 2, pp. 78–98.) In other cases we use a corrected *chi* square, $P(\chi^2)$.

to hit upon a satisfactory solution to their meal-time difficulties was highly frustrating to them, and their efforts to find a solution would lead them to generalize from other spheres of child care. The mothers who were rated high on general dominance in child-rearing should thus be those most likely to behave in a dominant or aggressive manner in the specific meal-time situation.[3] *Table II* shows that this was the case.

TABLE II *DEGREE OF GENERAL DOMINANCE, BY TYPE OF MOTHER*

Type	*High*	*Low*
Aggressive	23	5
Persuasive	5	12
		$P(\chi^2)<.01$

Most aggressive mothers also believed that their behavior had no undesired effect on their children's personalities, while the persuasive mothers

TABLE III *EXPECTED EFFECT OF AGGRESSIVE BEHAVIOR ON PERSONALITY, BY TYPE OF MOTHER*

Type	*No Effect*	*Undesired Effect*
Aggressive	21	7
Persuasive	5	12
		$P(\chi^2)<.01$

3. Three questions were asked concerning parental practices: What amount of explanation was offered to the child as a basis for parental demands? To what degree was the child required to fit into the routine of the home as established by the parents? and what degree of authoritarianism was present in the training of the child in control of elimination? Responses to the three questions were scored high, middle, or low. The possible responses were ranked from the distribution of high-high-high to low-low-low, and then a cut made at a rank so as to most equally divide the total number of subjects into two groups.

believed the opposite. The distribution on this characteristic is given in *Table III.*

In addition, twenty-five of the twenty-eight husbands of the aggressive mothers either approved of such behavior or offered no objections, so that these mothers were not deterred from its performance by expectations of punishment. The contrast with the persuasive mothers is shown in *Table IV.*

TABLE IV *REACTION OF HUSBAND TO AGGRESSIVE BEHAVIOR, BY TYPE OF MOTHER*

Type	*Approval or No Effect*	*Disapproval*
Aggressive	25	3
Persuasive	9	8
		$P(E)<.01$

The aggressive mothers, finally, were less strongly motivated to avoid the emotional disturbances caused in their children by such severe treatment, e.g., crying, withdrawal, tantrums, and the like.[4] See *Table V.*

TABLE V *DEGREE OF CONCERN OVER THE CHILD'S CONTENTMENT, BY TYPE OF MOTHER*

Type	*High*	*Low*
Aggressive	10	18
Persuasive	13	4
		$P(E)<.01$

Turning briefly to the twelve intermediate-type mothers, who threatened or yelled at their children but did nothing more than this, we find that they differ significantly from the persuasive mothers in one respect, and from the aggressive mothers in others. They differ from the persuasive mothers in that more of them show dominance in child care (and so are like the aggressive mothers). They differ from the aggressive mothers in that more of them believe that severe behavior reduces the child's eating, and also has undesirable effects on the child's personality. In addition, more are married to husbands who disapprove of severe behavior, and more of them are disturbed by the emotional upsets engendered in the child from such behavior.

4. The degree of concern was rated by the researcher on the basis of answers to three questions: How severely did the mother punish the child when she felt he was bad? What did she do when the child was upset or crying? and What did she do when other children were disturbing her child by physical or verbal aggression? Responses to these three questions were treated in the same manner as responses to the three questions concerning dominance.

In sum, the aggressive mothers have a greater tendency toward severe behavior, and are less restricted in their expression of it by other concerns. The intermediate mothers have a greater tendency toward such behavior, but are restricted in its expression so that it does not emerge in such full-blown fashion as with the aggressive mothers. Finally, the persuasive mothers have less tendency toward severe behavior, and at the same time are or would be restricted in the expression of such behavior.

TRIAL AND ADOPTION OF PERMISSIVE BEHAVIOR

In the following pages we turn our attention to the process of change in behavior. We shall consider change as taking place over a temporal span, running from an initial exposure to an innovation toward some end point at which "final" acceptance or rejection is said to have occurred.

In some instances of change it is necessary only to overcome initial resistance and to induce subjects to make a trial of the novel behavior, because the satisfying results are so readily apparent that the new behavior becomes substituted for the old. In such instances there is some merit in treating change only in terms of an end product. However, in most situations involving social behavior the satisfying consequences of the novel pattern do not occur so quickly, and it is necessary for the subjects to be sustained through a period during which there is no immediate reward. In such situations it is of value to view change as a temporal process, because different variables may influence different phases of the process.

In the case of permissive behavior in feeding, the characteristic initial result of performance is to *decrease* the amount eaten by the child. With customary pressures suddenly removed the child usually eats very little for a number of days. The satisfying result of an increase in eating is delayed, dependent upon a two to three week period of sustained performance of the pattern. Thus it was necessary in the cases at hand that the subjects not only made initial trials of the pattern, but that they continued to employ it to the point where the actual rewarding results, in the form of their children's improved eating, emerged to stabilize their adoption of the new practice.

We will try to capture the process of change by contrasting three groupings of subjects, distinguished by differences in their participation in certain phases of the process. The three groups considered are the following: the mothers making no trials, the mothers making trials but not continuing performance, and the mothers who adopted the pattern. *Twenty-six* subjects were in the non-trial group, having made no performance of permissiveness at all during the period between the two interviews. *Sixteen* cases were in the rejecting-after-trial group, being mothers who had initiated performance of the pattern but had not continued it. The remaining *eight* subjects, comprising the adopting group, were those mothers who substituted permissive

behavior for their previous mode of action, and, at the time of the re-interview, were meeting all food refusals with a permissive response.

All the subjects who made trials made them within a day or two after the initial interview. The mothers who adopted the pattern continued their performance from the initial trial through a mean period of fifteen days before their children began to eat better. The sixteen mothers who rejected the pattern in the end made shorter trial performances, four of them continuing for approximately one week and the other twelve making intermittent trials at different meals for a short period of time after the interview.

The use of these three groupings defines two problem areas. The first concerns the factors which generated *trials* of permissiveness by the eight adopters and the sixteen others who also made trials, in contrast to the twenty-six who did not. The second problem concerns the factors which differentiated the eight *adopters* from the sixteen others in the trial group, so that the former made a sustained performance and went on to accept the pattern as a standard response.

The logic of our inquiry defines the nature of the inter-group comparisons to be made. First, when we seek to understand the importance of the factors for the initiating of trials, we must compare the total trial group, composed of the eight adopting cases and the sixteen who tried but rejected the pattern, with the non-trial group of twenty-six mothers. Secondly, when we give our attention to the influence of the factors on the adoption process itself, logic calls for a comparison of the eight adopting cases with the sixteen cases who rejected the pattern after trial.

This logic, however, assumes that there are two discrete sets of factors, one set restricted to engendering trials, the other to sustaining performance and issuing in later adoption. This is not necessarily the case in phenomena of change. Certain of the factors are operative throughout the process of change, affecting both trial and adoption, their influence being overridden at the outset for some of those making trials, but arising during the course of the trial period to cause some of the subjects to drop the new behavior. But this type of influence is only made clear by a third type of comparison, that between the adopting group and the total non-adopting group, whether trials were made or not. That is, where the influence of some factor is spread out over the time span, touching some early and thus dissuading them from trials, and others later, thus leading to discontinuance of performance, its total effect upon change can only be captured by a comparison of adopters with the non-adopting group taken as a whole.

Therefore we will speak of three types of factors relevant to change in behavior: those affecting initial performance; those affecting adoption among those subjects making trials; and those factors influential in both instances.

At the time of the first interview all of the mothers had certain customary behavior which they displayed when their children refused to eat. Moreover, all of them believed at that time that performance of permissive

behavior would have the consequence of *decreasing* the amount which their children would eat, either immediately, or later, or both. We may therefore examine the impact which jarred nearly half of them out of performance of their customary behavior into performance of a pattern which they believed to be unrewarding—i.e., the initial factor which started the process of making trials.

The burden must be laid upon the interview experience itself, in which the interviewer played the role of a doctor. It was not the fact that the mothers were presented with a way of behaving which they had not conceived of before, and had therefore never employed. All knew of the pattern of permissiveness as a potential alternative, although they had never used it. Rather, permissive behavior was re-emphasized in the close interpersonal situation created by the interview with the "doctor", so that the merits of the pattern were reconsidered.

If this is the case, then one may ask why all of the mothers did not make trials, since all went through the interview experience. At this point one can assume a difference in attitudes toward doctors, and in evaluations of advice received from such a professional source.

The first factor we wish to consider then is the prestige accorded to doctors by the mothers. The mothers were asked to give a general rating of previous advice received from doctors about child care, in terms of how "helpful" it had been. We assume that the higher this rating, the more influential the current advice of the doctor would be.

The distributions in *Table VI* can be utilized in making our three inter-

TABLE VI *CHANGE BY PRESTIGE RATING OF DOCTORS*

	High	*Middle or Low*
Adopted	6	2
Rejected after Trial	12	4
Non-trial	7	19

group comparisons. The comparison of the adopting with the rejecting-after-trial group shows no significant difference. However, the difference between these two combined, i.e., the total trial group, and the non-trial group is significant at $P(\chi^2)$ less than ·01. Finally, the comparison between the adopting group and the other two, i.e., the total non-adopting group, shows no significant difference. How is this information to be interpreted?

The first comparison indicates that the doctor's prestige does not affect adoption itself among those making trials. The second indicates that, on the other hand, this factor is highly important in regard to engendering trials or initial performance. The third comparison indicates that, even when opportunity is given for the possible effect of this factor upon adoption over

the whole time period to appear, there is still no significant relationship. Therefore we can conclude that the prestige rating of doctors was influential in initiating performance, but not upon adoption itself.

It is especially noteworthy that if our analysis had viewed change simply as an end product, and compared only categories of changers and non-changers, we would have had only the third comparison made above. This comparison does not show the influence of the prestige rating of doctors, and in such a limited analysis the research would have led to the erroneous conclusion that this factor was not important in the change. Yet it is clear, as a result of the sequential analysis made here, that the factor was highly important in stimulating trials of the new behavior, and, since there can be no change without a first performance, was also important for the net resultant change itself.

In addition to the advice of the doctor, a second factor also influenced expectations sufficiently to bring about some initial trials of the novel pattern. This factor was the reason, or underlying motive, for which the mothers' wished their children to eat. We have up to now treated the concern over eating at an undifferentiated conceptual level. This motive, however, as is the case with all, can be viewed as an instrumentality of some more basic concern.

During the first interview additional data had been collected as to "why" mothers wished their children to eat. Three reasons were given by the mothers: 1. to have the children show appreciation of their efforts in preparing the food; 2. so that the children would be bigger; and 3. so that the children would enjoy a good diet. Four mothers gave the first answer, twenty-four the second, and twenty-two the third.

This difference between the mothers' reasons for wishing their children to eat better was not related to their customary behavior. It is important, however, in regard to change toward more permissive behavior, because the rewards to be gained from the results of permissiveness differ depending on the type of underlying motive which the mother possesses. To leave the analysis at the more abstract conceptual level employed up to now would be to omit completely the influence upon change of the underlying motivational component present in the mealtime situation, and also the fact that the behavior change was in part a result of the differential rewards expected from the novel pattern.

The recommendation of permissive behavior is based on the idea that a child will, if left alone, eat the correct amounts of the correct foods in terms of his bodily requirements. It is the mother concerned over diet who would be most fully satisfied by allowing a child to eat in accord with his needs. The other two concerns—size and appreciation—involve parental standards to the extent that, even if the child ate what he needed, this still would not necessarily satisfy the parent. For example, with reference to concern over the child's size, the problem children as a group were already very close to

the average. It is not likely, then, that their physiological requirements would lead them to eat enough more to satisfy their mothers. Even though permissive treatment may increase the child's eating up to the point of his own needs, this increase may be relatively non-satisfying to mothers who are concerned over things other than the child's diet.

It may be recalled that during the interview it was explicitly pointed out to each subject that permissive behavior customarily results in the child's selecting an adequate diet. Nothing was said in regard to weight gain or, of course, show of appreciation. This clear statement of the function of permissive behavior with regard to an adequate diet, together with the fact that most of the mothers concerned over weight already knew that their children were of approximately normal weight, is the basis for the inference that the subjects made different evaluations of the probable gain from permissive behavior in terms of their underlying reasons for wanting the children to eat. Mothers concerned over diet could reasonably expect more satisfaction than mothers concerned over weight gain or appreciation.

The relationship between the three groups and the type of motive is given in *Table VII.* The distribution shows a relationship slightly different

TABLE VII *CHANGE BY MOTIVE FOR WANTING THE CHILD TO EAT MORE*

	Diet	*Size and Appreciation*
Adopted	7	1
Rejected after Trial	8	8
Non-trial	7	19

from that of the preceding variable. The first comparison, of the adopting with the rejecting-after-trial group, shows no significant difference. This suggests that the motive factor does not affect adoption. The second comparison, of these two groups taken together with the non-trial group, gives a difference significant at $P(\chi^2)$ less than ·05. The indication is therefore that this factor is influential with regard to the subjects making initial trials of the new behavior.

The third comparison, however, between the adopting group and the other two non-adopting groups, also results in a difference significant at P(E) less than ·05. We must conclude for this factor, therefore, that it is influential upon the adoption aspect of change, although its effect is accentuated by being spread over the whole process. The major portion of its influence lies in engendering trials, but it is nevertheless of importance also in the continuation of the trials.

Two more variables affected expectations about the consequences of permissiveness in regard to the child's eating. In both of these it will be shown

that influence was spread over the temporal process of change, affecting not only the initiation of trials but the continuation of these trials subsequent to adoption.

The first of these is the amount of general dominance which the mother characteristically used in her child-rearing. This factor, which had operated through generalization to lead the aggressive-type mothers to severe behavior, would, in the same manner, influence estimates of the results to be obtained from permissiveness. Given the recommendation of the permissive pattern, it is to be expected that the mothers lenient in child-rearing would have more favorable expectations for the extension of lenient treatment into the area of feeding. *Table VIII* gives the distribution on dominance by the three

TABLE VIII *CHANGE BY DEGREE OF GENERAL DOMINANCE*

	High	*Low*
Adopted	2	6
Rejected after Trial	10	6
Non-trial	20	6

groups. Here the comparison between adopting and rejecting-after-trial groups is not significant; nor is the comparison between these two and the non-trial groups. The first indication is that this factor is not relevant to change, distinguishing neither between adopters and rejecters, nor between trials and non-trials. However, the third comparison, between adopters and all non-adopters, is significant at P(E) less than ·05, thereby indicating that the influence of this factor serves to retard change, but that the influence is spread out over the change process, affecting both trials and subsequent adoption; thus, the importance of this factor can be assessed only when its influence on non-adopters is summed and contrasted with that on adopted. We may conclude, then, that degree of child-rearing dominance was influential upon the mothers' expectations about permissive behavior.

The second additional factor affecting the subjects' expectations about the novel pattern is the sources of information, other than doctors, which they rely upon in their child-rearing practices. One would expect the sources of information tapped by mothers to be significantly related to their tendency to change. One mother may have read the same thing in different books; another may talk daily with friends and get different advice; one may be constantly reassured, another constantly told she is foolish. The effect of the variety of information received would vary with the nature of the indications, the frequency with which they were given, and the subjects' beliefs about their validity.

The subjects were rated as being exposed to advice either favorable or

unfavorable to permissive behavior, i.e., advice which clearly asserted that such behavior would increase or decrease the amount the child ate. The sources used in this rating included only those on which the subjects stated they had a high reliance, and which they referred to weekly or more often. *Table IX* shows the distribution of the three groups in regard to the favorableness or unfavorableness of the advice received.

TABLE IX *CHANGE BY CUSTOMARY EXPOSURE TO SOURCES OF ADVICE CONCERNING PERMISSIVENESS*

	Recommend Permissiveness	*Recommend Other Method of Feeding*
Adopted	8	0
Rejected after Trial	7	9
Non-trial	8	18

The comparison of the distribution of the adopting and the rejecting-after-trial groups shows a difference which is significant at P(E) less than ·01. Comparison of these two with the non-trial group similarly shows a significant difference at P(χ^2) less than ·05. Already one can draw the conclusion that this factor influenced adoption among those making trials, and in addition was influential in regard to the undertaking of trials themselves. Therefore the third comparison, between the adopting and total non-adopting group, must also indicate the sizeable influence of this factor upon both aspects of change, an influence so large that it appears in the more specific between-group comparisons as well as in this overview. The comparison does in fact show a difference that is significant at P(E) less than ·01. In sum, significantly more of the mothers not changing were exposed to advice consistently suggesting that some other feeding method was preferable, and this factor was influential throughout the whole process.

The foregoing four factors have been considered in their influence upon expectations relative to the relationship of permissiveness to the child's eating. In addition to this analysis, one must also inquire into the expected effects of permissiveness with regard to the three remaining motives of the mothers: those concerning social approval, the child's contentment, and his character.

Two of these motives, for the development of an adequate personality in the child, and for his current emotional happiness, may be discussed together. For the mothers who were of the aggressive type, permissive behavior had potential rewards which their customary severe behavior did not carry. For mothers of this type who, first, expected that their behavior was a deterrent to the children's personality development, and second, were more highly motivated to keep their children emotionally content, per-

missive behavior could potentially bring increased satisfaction in this respect. This would not be equally true for the remainder of the aggressive subjects who were not so concerned over these matters. Neither would it be true for mothers of the persuasive and intermediate types, because their earlier behavior was not so detrimental to the satisfaction of the two motives, and who therefore had less to gain.

It would be expected, therefore, that mothers who displayed aggressive behavior in spite of their expectation of bad results, and in spite of their desires to keep their children happy, would be more likely to change toward permissive behavior than would other aggressive mothers.

This turned out to be untrue with reference to concern over the child's personality. Comparison between those aggressive mothers who thought that their customary behavior had no effect on the child's personality, and those who thought that it did, showed no difference in tendency to try the novel behavior. We can suggest that the subjects did not sufficiently perceive the relationship between the novel behavior and its consequences for personality for this expectation to influence their behavior.

The connection between behavior and consequences is more readily made in the case of the child's contentment. It requires little discernment by the mothers to realize that beating a child upsets him, and abstention from punishment makes him feel better. *Table X* gives the distribution for the

TABLE X *CHANGE BY DEGREE OF CONCERN OVER THE CHILD'S CONTENTMENT* * *(AGGRESSIVE-TYPE MOTHERS ONLY)*

	High	*Low*
Adopted	0	1
Rejected after Trial	6	3
Non-trial	3	13

* Two of the original total of 28 aggressive-type mothers were among the 7 cases not re-interviewed.

aggressive mothers only, within the three groups, by the strength of motive for the child's contentment. The small frequencies permit us to make only the comparison between the trial group (adopting and rejecting-after-trial) and the non-trial group. The difference between them is significant at P(E) less than ·05. We may conclude, therefore, that this factor was influential in regard to the making of initial trials, but we can say nothing about its possible additional affect upon adoption itself.

Next we may consider the influence of the expected consequences of permissive behavior in regard to social approval. The group context into which the novel behavior was introduced was that of the family, and it would be expected that the favorable or unfavorable reactions of the

husbands would have an effect upon the continuation of the wives' performance. However, the normative situation with respect to the permissive feeding practice was so unstructured that a different thing occurred. What appears to have happened is that the husbands' reactions to the novel pattern developed in conjunction with the wives' performance of it, rather than influencing the performance by virtue of their prior existence.

Table XI shows the *wives' reports* of the reactions of the husbands at three

TABLE XI *CHANGE BY REACTION OF HUSBAND TO PERMISSIVE BEHAVIOR, AT THREE TIME INTERVALS*

		Adopted	*Rejected After Trial*	*Non-Trial*
First Interview	Disapproval	1	6	7
	No Reaction	7	10	17
	Approval	0	0	0
Start of Trial Period	Disapproval	2	5	12
	No Reaction	2	7	10
	Approval	4	4	4
At Re-interview	Disapproval	0	9	12
	No Reaction	0	4	10
	Approval	8	3	4

time intervals: the first interview, the start of the trial period, and the re-interview. These data may include a large measure of distortion by the wives, and therefore the interpretation which follows warrants little emphasis.

In this case a slightly different mode of analysis is required, in that the factor being considered did not itself remain constant during the change period. To assess its influence upon the initiating of trials, it is necessary to consider the distribution between trial and non-trial groups at the start of the trial period. No significant difference lies in this comparison, and hence we are justified in stating that this did not influence trials.

When we turn to assess the influence of the husbands' norms upon adoption, it seems desirable to take the distribution which obtains at the time of the re-interview, and to compare the adopting and the rejecting-after-trial groups. This results in a difference significant at P(E) less than ·02, for the four-fold tables of disapproval versus approval, disapproval and no reaction versus approval, and disapproval versus no reaction and approval. We may tentatively conclude that this factor was influential in adoption, for those subjects making trials.

The final general factor to be considered is the actual result of performance of the novel behavior pattern. The results must almost certainly be an

improvement in the child's eating, if the mother is to commit herself to the pattern as her customary response. Despite recommendations by doctors, ideas gained from newspapers, intimations of friends, reactions of husbands, and the like, it is necessary that the child begin to eat better if the novel pattern is to become completely adopted. The full consequences of continued performance must themselves take over the influential role and give certainty to the subjects' expectations.

The eight mothers who adopted permissiveness were the only ones who had made a sustained trial for a period of at least two weeks. The children of all of these began to improve in eating after a mean period of fifteen days, and it was this improvement which served to fix the adoption of the new pattern. In contrast, the children of the other two groups showed little or no improvement. *Table XII* contrasts the three main groups considered in the

TABLE XII *CHANGE BY FREQUENCY OF CURRENT REFUSALS*

	Daily	*Less than Daily*
Adopted	0	8
Rejected after Trial	10	6
Non-Trial	20	6

analysis on the frequency of food refusals at the time of the re-interview. At the time of the first interview all children daily refused to eat. Any current frequency less than daily thus represents an improvement in the child's eating. The difference between the adopting and the other two groups is significant at P(E) less than ·01 in both instances.

The differences shown indicate that the group of women adopting the pattern had results far more satisfying than did the other two groups. Lack of improvement for the latter came either from the fact that no trials at all were made, or that those making trials made intermittent, rather than sustained performances, and soon regressed to non-permissive behavior.[5]

SUMMARY AND CONCLUSIONS

We have considered the process of acceptance of new behavior, namely, the permissive feeding of children, by a group of mothers. Our consideration involved two phases. First: the analysis of the foundations of the already-existing behavior patterns of the mothers, with which the innovation

5. The subject children of the study were originally selected within the age range of 18 to 36 months to minimize the possibility of an improvement in appetite from physiological maturation during the course of the study. Gesell and Ilg (4, pp. 238–242 and 259–261), as well as other writers, state that during the age range involved there is little improvement by "poor eaters" from maturation. The improvement can safely be laid to the mother's behavior change in the mealtime situation. The twelve mothers who did not adopt the pattern but nevertheless reported a decrease in food refusal 1. reported less of a decrease, and 2 had themselves changed in behaviour *toward* permissiveness, i.e., changed to a less severe pattern, without adopting the recommended pattern.

competed. Second: the assessment of the influence of a number of factors upon the various steps in the process of acceptance, from initial exposure and early trials through continued trials and final commitment. We have suggested that acceptance requires an initial favorable change in expectations regarding the consequences of the novel behavior, and that these expectations must be sustained through a period of trial performance to the point at which the actual satisfying results of continued performance occur to stabilize the adoption.

Significant relationships between a number of factors and behavior change have been set forth, although our observations do not permit us to evaluate the relative influence of each of these factors. These were: the prestige rating of doctors; the relative satisfaction to be gained from the change; the degree of dominance characterizing the mother's child-rearing; the exposure to favorable or unfavorable advice concerning the change; the encouragement and approval received from husbands; and, last, the actual results obtained from employment of the new behavior pattern.

In conclusion, we speculate briefly on three implications of the findings. First, the study of the relative satisfaction to be gained from the change demonstrated the need for care in the analysis of motives present in the subjects. The evidence suggests that to overlook this, in analysis of change, may result in erroneous estimates of the rewards or functional adequacy of the novel pattern, and of its probable adoption when introduced as an alternative mode of behavior.

The two further points are specifically relevant to the type of change studied here, in which the rewards of change are delayed.

One is that, although change was influenced by the amount of prestige accorded to doctors, this influence acted chiefly upon initial trials of the pattern, and was not important in its continued performance. This supports the idea that in instances of behavior change where this factor alone is present, change will take the form of many trials but little net result.

The other and last point concerns the proposition that, in undertaking novel behavior for which the satisfactions are delayed, it is necessary that the performance be supported and sustained in some manner. The evidence concerning the influence of the sources of information drawn upon in the trial period, together with the influence of the husband's reactions, suggests that one important source of support is a favorable group context for the individual. This is important for two reasons. In such a context the individual's expectations of future satisfaction are reinforced and sustained by those about him. And in such a context, moreover, the individual is gaining, during the period when results have not yet occurred, satisfaction of his basic desire for security and approval, through encouragement and other expressions of solidarity. Thus groups not only sustain the individual's expectations, but reward him at the same time.

BIBLIOGRAPHY

1. ALDRICH, C. A., and ALDRICH, MARY M. *Feeding our Old Fashioned Children*. New York: Macmillan, 1941.
2. FISHER, R. A. *Statistical Methods for Research Workers*. New York: Harper, 1950.
3. FRENCH, T. M. "Goal, Mechanism and Integrative Field", *Psychosomatic Medicine*, 1941, *3*, 226–252.
4. GESELL, A., and ILG, FRANCES L. *Feeding Behavior of Infants*. Philadelphia: Lippincott, 1937.
5. HILLIARD, A. L. *The Forms of Value; the Extension of a Hedonistic Axiology*. New York: Columbia University Press, 1950.
6. HULL, C. "Goal Attraction and Directing Ideas Conceived as Habit Phenomena", *Psychological Review*, 1931, XXXVIII, 487–506.
7. —— "The Mechanism of the Assembly of Behavior Segments in Novel Combinations Suitable for Problem Solution", *Psychological Review*, 1935, XLII, 219–245.
8. MOHR, G. J. "Emotional Factors in Nutrition Work with Children", *Mental Hygiene*, 1928, *12*, 366–377.
9. PARSONS, T. *The Social System*. Glencoe, Ill.: Free Press, 1951, London: Tavistock Publications, 1952.
10. RAPAPORT, D. "The Conceptual Model of Psychoanalysis", *Journal of Personality*, 1951, *20*, 56–81.
11. SPOCK, B. *The Pocket Book of Baby and Child Care*. New York: Pocket Books, 1946.
12. STOUFFER, S. A., GUTTMAN, L., SUCHMAN, E. A., LAZARSFEID, P. F., STAR, SHIRLEY, A., and CLAUSEN, J. A. *Measurement and Prediction*. Princeton: Princeton University Press, 1950.
13. TOLMAN, E. C. and BRUNSWIK, E. "The Organism and the Causal Texture of the Environment". *Psychological Review*, 1935, *42*, 43–77.
14. TROLAND, L. T. *The Fundamentals of Human Motivation*. New York: D. Van Nostrand, 1928.
15. *Your Child from One to Six*. Children's Bureau. Publication 30. Washington: Federal Security Agency, 1945.

BIOGRAPHICAL NOTE

ORVILLE G. BRIM Jr. received his M.A. and Ph.D. degrees in sociology from Yale University in 1949 and 1951. He was research assistant in the study of Cultural Factors and Talent Development at Yale during 1951 and 1952. He currently is an Instructor in the Department of Sociology and Anthropology of the University of Wisconsin.

FOUR FORMULAE FOR SELECTING LEADERS ON THE BASIS OF PERSONALITY

RAYMOND B. CATTELL AND GLEN F. STICE

I. TWO CONCEPTS OF LEADERSHIP AND LEADERSHIP MEASUREMENT

Although the literature on leadership measurement and selection is considerable (1, 12, 13, 14, 15, 16, 17) it cannot be denied that psychologists and administrators charged with selecting people to fill pre-established positions of leadership are currently disappointed with the results. As a consequence, the psychologists are turning toward other conceptualizations, or are emphasizing situational determinants, as in the work of Gibb (12), while the selection of leaders continues to be based largely upon such things as school marks, experience, and other vaguely defined rules of thumb.

The failure of research to lead to either good theoretical generalizations or to effective applied work arises, in the opinion of the present writers, from the lack of meaningfulness, validity, and relevance *in the personality measurements themselves*; from failure to use sound operational definitions of a leader, and perhaps from a tendency to ignore low, but sometimes suggestive, relationships where they have been found. However, the problem now seems to be ripe for more fruitful investigation; first, because the advance of personality research has made it possible to use measures of meaningful, factorially-independent, personality traits where previously personality has been measured only by *ad hoc* tests of unknown composition and validity; and second, by virtue of the growth of new concepts, as a result of broader experimental work, in group dynamics.

The recent theoretical developments and empirical test results in regard to the definition of unitary personality traits seem to be well known (5, 18). It is necessary therefore to indicate advances only in the second matter, namely, the revised concept of leadership measurement in the group dynamics setting. In a theoretical treatment we have suggested that there are always two distinct areas from which leadership measurements can be

obtained: 1. measures of the internal group process, and 2. measures of the total group performance or product. The former, by far the most common basis in past research reports, includes direct observations of the kind of interaction and its channels, and also, less satisfactorily, retrospective and introspective sociometric reports. By contrast the "group performance" criterion of leadership, perhaps more frequently used in the hard realm of international politics, asks the question, "How well did this group perform under this leader?", and takes as a measure of leadership adequacy, a measure of group performances—usually corrected in practice by intuitive allowances for situational factors which are believed to be important.

The reluctance of some leadership researchers to shift to this second, "behavioristic" definition of leadership seems to be based largely on a theoretical or practical inability to separate the effectiveness of the group population from the effectiveness of the leader, as they contribute to the given performance. This difficulty vanishes completely in theory and largely in practice as we develop methods of measuring the *syntality* of the group with reference to which the leadership occurs.

If the characteristics of a group are measured along independent dimensions, as suggested in a recent research (2, 9), we obtain a profile of what may be called the group syntality—the dimensions of group behavior. This measurement of syntality permits a precise statement—in terms of change on syntality dimensions—of the way in which a group changes under the impact of a new situation, internal organization, or leader. It is proposed then that the proper "behavioristic" definition of a leader should be in terms of *the extent of his influence upon each of the relevant dimensions of the group syntality* (7).

With regard to the *designation* as apart from the *measurement* of leadership, there appear, correspondingly, to be two criteria, not always consciously distinguished. On the one hand the leader is taken as the person who occupies a given status in the group. He may occupy this either through election by the group or appointment from outside the group, but in either case he is presumed to be the person who performs certain acts including, generally, giving directions to the group and serving as its spokesmen. On the other hand the leader is taken as the person, whatever his formal status, whose behavior has significant influence upon what the group does or how it does it. Too often, it would seem, studies of leadership have failed to recognize this distinction clearly. Much of the mere consistency (as distinct from validity) and some of the resulting prestige enjoyed by sociometric ratings of leadership resides circularly in a single simple fact: while group members can agree upon which members hold the position of "leader", the inconsistencies between such agreement and the attempts to evaluate leadership in terms of group productivity is evidently due to a low correlation between actually influential behavior and formal leadership status.

In the present paper it is proposed to examine leadership in a way which

has not been done before, namely, by considering it in the content of many diverse measures of group performance, resolved into syntality dimensions, This was made possible by an Office of Naval Research study[1] on 100 groups of 10 men each, the results of which, as far as the group measurements are concerned, have already been presented in more detail elsewhere (8, 9).

The findings to be described here are those based, on the one hand, on the measurement of individuals in terms of well-defined personality factors and, on the other, upon defined groups performing in standardized situations. In this way it is possible to obtain a number of measurements of leadership with a degree of uniformity and objectivity which has perhaps seldom been possible in the past. The theoretical analysis of leadership *designation* and *measurement* along the lines indicated above, has been set out in detail elsewhere (6, 7). Here we shall deal with a particular experiment exemplifying the principles there discussed.

II. DESIGN OF THE EXPERIMENT

It was our aim to take a moderate size face-to-face group concerned with a variety of activities typical of those which such groups are commonly called upon to face. A greater degree of control over the selection of the populations and over the experimental conditions of group behavior was exercised than has generally been possible in leadership observations. Thus the populations were composed of persons of the same sex (men), of a fairly narrow age range (18–30), and chosen as having no previous acquaintance of one another in such group life. Each of the 80 groups (20 of the 100 were dropped because of incomplete data) was in competition with the other groups for a considerable monetary prize ($100.00 to be given to the best group in each set of 10 groups). After the testing, by the personality tests, of the individual members who were to be in the group, each group met for three sessions of three hours each. During these sessions it was called upon to perform a wide variety of activities, such as a construction task, a committee meeting, a jury-like decision-reaching situation, a code problem-solving situation, a discussion—designed to arrive at the common attitude and interest of the group on certain questions—and so on. The forty-four distinct performance measurements so obtained have been described fully elsewhere (8, 9). The measures when intercorrelated and factor-analyzed yielded about 16 group dimensions—13 of them sufficiently stable for measurement—on each of which each group had a measurable value.

The observations made on the group during its performances covered not

1. The collection of the data, upon which the analysis described in this paper was done, was accomplished under Contract NR 172–369 sponsored by the Human Relations Branch, Office of Naval Research. The authors wish to thank the ONR for this support and to thank the Commanding Officers, USAF, OCS, Lackland AFB, Texas and USN, Recruit Training Command Great Lakes, Illinois, who provided the subjects and the space for testing them.

only the measurements of time and errors in the various total group activities, but also observations by two observers in a great variety of interactions within the group, sociometric reports by members of the group on their relations with other members, and records of who had been the designated (elected) leaders, together with the circumstances under which they were elected. Owing to a shortage of computational resources it has not yet been possible to classify the groups according to their standing on the 13 syntality dimensions. Thus, no direct measures of leadership in terms of influence on these dimensions are available, though the observers' rating and interaction data on the influence exerted by an individual on the group are available to be used as a short cut to the same principle. These observers, working under standardized, repeated conditions, observed the effects of the various members of the group upon the behavior and performance of the group, and they recorded each group-influencing act; marking it with the number of the person who made it. Secondly, after the first session of three hours, in which no formal leader was specified, the group were invited to select a leader from among their members. In every case they chose to do so. Three times during the course of the second meeting, and again at the end of the meeting, they were asked to hold elections to decide who would serve as leader for the next activities. In all cases they were permitted to re-elect their contemporary leader or to select a new one.

These represent two extremes, as it were, of the possible leadership observation. But the full rostra of experimentally independent leadership evaluations, made in the course of these experiments, admit intermediate grades. Those to be reported in this paper are:

1. "*Problem-Solving Leader*" (252)[2]

Immediately after each situation each observer (there were generally two) checked, in a mimeographed form, the code name for each group member who, at any time in the course of the situation, had been observed to have had a significant influence upon the group. The total number of checks (44 was the largest possible number) was computed for each member, and, finally, the members were ranked in the group in terms of the number of checks they received. Those with the most checks were ranked high in amount of leadership shown.

2. "*Salient Leader*" (253)

To be scored a leader in this category two observers had independently, and for the same situation, to star the subject as being the *most important* leader in at least one of the twenty-two situations presented in the course of the three meetings. All subjects upon whom two observers had agreed at least once were taken as leaders. Since only one observer was used for the Great Lakes groups this measure could not be computed for them.

2. The number in parentheses following each title refers to the index number of the measurement in the master list of variables (9, pp. 20–39) from which the score was computed.

3. "*Sociometric or Popular Leader*" (112)

At the end of the final meeting all group members were asked to recall which of them had shown leadership in the course of the group meetings and to record their numbers in the order of importance of their leadership. Six blanks were provided.

In computing "leadership scores" by this criteria, the first blank was weighted 4 points, the second 3, the third 2, and all remaining ones 1. The distribution of scores was then broken into "leaders" and "non-leaders" at what appeared to be the point of inflection between two modes in this distribution.

4. "*Elected Leader*" (413)

Any person who was elected leader on one or more occasions was taken as a "leader" using this criteria.

To complete the description of the experimental work it is necessary only to describe how the individual personality measurements were made. Since the measurements could be made on co-operative subjects, and could be made before the creation of the group situations in which they might become rivals for leadership, it seemed reasonably safe to depend upon a pencil and paper questionnaire, especially because no behavorial measures of the primary personality factors were available at the time the work began in 1947. Each subject was accordingly given both the A and the B forms of the *16 Personality Factor Questionnaire* (4). This questionnaire is based upon a large scale factorization of items chosen to represent a wide range of personality responses and many areas of behavior (4, 5). It has been developed by various item-analytic purification technics and now measures 16 factors by either 20 or 26 questions each. These factors have been correlated with ratings and also with a number of other tests, such as the Humor Test (10), and the Music Preference Test (11). The factors, which are symbolized by letters of the alphabet, represent such dimensions as emotional maturity versus neuroticism, cyclothymia versus schizothymia, surgency versus desurgency, high intelligence versus low intelligence, dominance versus submissiveness, and so on. Although these factors are practically orthogonal in the original factorization, they have some moderate intercorrelations, ranging from ·0 through +·4 in the questions by which they are now measured. The fact that this test was given to subjects in a slightly abbreviated form, compared with the item total in the present standard test, creates some slight difficulties in equating the raw scores with the norms for the standard published 16 P.F., but these difficulties can be overcome by suitable corrections, and in any case they do not affect the main findings. The subjects were exhorted to respond frankly and were assured that the responses would be kept private and would not influence their occupational or student standing.

III. THE EXPERIMENTAL POPULATION

Although 80 groups, involving 800 men, were used for the principal study, the present analysis of leadership behavior was worked out in only 34 of these groups, including a total of 334 men. The discrepancy between the number of men and of groups was caused by a few cases where the personality questionnaire was not filled out. Fourteen of the groups were composed of subjects who had just arrived (within 72 hours) at San Antonio, Texas, to begin Officer Candidate School in the Air Force, and twenty of the groups were of subjects who had arrived (within 12 hours) to begin Boot Training at Great Lakes, Illinois.

IV. ANALYSIS OF THE RESULTS

The plan of analysis is a very simple one. First we will compare the profiles of leaders with non-leaders, where leaders are chosen separately for each of the criterion described above. In this way we hope to discover which of the measured personality factors are significant, how far each is significant for leadership, and how it may be possible to weigh the significant factors in order to produce a pattern measurement distinguishing leaders from non-leaders.

We will then compare the four profiles of leaders in order to discover what if any relation exists between the method of selection and the personality characteristics of the resulting selectees.

The results in terms of mean scores for all criteria of leadership are shown in *Table I*, the actual critical ratios for differences that were found to be significant in *Table II*, and the size of the differences in terms of standard score units are plotted on *Figure 1*.

It will be seen at once that a number of differences exist which are significant at the 5 per cent (C.R. = 1·96) and at the 1 per cent (C.R. = 2·58) levels of confidence, and that in most cases there is a *tendency* for the *same* personality factors to show differences for each of the four different criteria of leadership. Personality factors showing differences in the same direction for all four criteria are:

C Emotional Maturity or Ego Strength
E Dominance
G Character Integration or Super Ego Strength
H Adventurous Cyclothymia
N Polished Fastidiousness
O (—) Absence of Worrying Anxiousness (Free Anxiety)
Q_3 Deliberate Will Control
Q_4 (—) Absence of Nervous Tension (Somatic Anxiety)

TABLE I — MEAN 16 P.F. SCORES FOR LEADERS AND NON-LEADERS

	Criterion 1: "Problem Solvers"				Criterion 2: "Salient Leaders"			Criterion 3: "Sociometric Leaders"			Criterion 4: "Elected Leaders"		
	High	Middle	Low	d	L	NL	d	L	NL	d	L	NL	d
A	17·8	17·9	17·5	·3	18·3	19·7	−1·4	17·1	17·9	−·8	18·6	17·4	1·2
B	11·4	10·4	9·7	1·7	12·5	11·7	·8	10·4	10·5	−·1	10·9	10·3	·6
C	35·1	34·1	33·6	1·5	36·1	34·2	1·9	34·6	32·9	1·7	34·8	34·0	·8
E	25·6	24·4	24·0	1·6	27·0	26·0	1·0	26·8	24·5	2·3	25·2	24·4	·8
F	26·9	25·6	25·0	1·9	26·1	26·2	−·1	25·6	25·9	−·3	26·4	22·5	3·9
G	22·1	21·0	20·6	1·5	22·4	21·3	1·1	21·7	21·1	·6	22·4	20·8	1·6
H	36·7	33·3	32·9	3·8	39·8	35·8	4·0	36·2	34·1	2·1	36·6	32·2	3·4
I	9·8	10·7	10·7	−·9	9·8	10·6	−·8	10·4	10·4	·0	10·0	10·6	−·6
L	18·1	18·6	18·4	−·3	16·8	17·8	−1·0	17·3	18·5	−1·2	17·9	18·6	−·7
M	19·0	20·1	20·1	−1·1	18·6	19·8	−·8	19·7	19·7	·0	18·6	20·2	−1·6
N	21·9	20·4	20·6	1·3	23·0	22·2	·8	22·1	20·8	1·3	21·4	30·7	·7
O	13·3	15·4	16·2	−2·9	11·3	13·6	−2·3	13·9	15·0	−1·1	13·0	15·9	−3·9
Q_1	19·9	20·3	18·8	1·1	21·5	21·2	·3	20·4	19·7	·6	19·7	19·8	·1
Q_2	14·1	14·1	14·7	−·6	15·9	14·5	1·4	15·7	14·0	1·7	13·6	14·6	−1·0
Q_3	25·8	24·3	23·2	2·6	26·7	24·6	2·1	25·8	24·3	1·5	26·0	23·6	2·4
Q_4	14·7	16·3	16·4	−1·7	12·0	15·0	−3·0	14·6	16·0	−1·4	14·7	16·0	−1·3
No. of cases	90	140	90		43	100		39	271		92	233	

* d's significant at the 5 per cent level are underlined.
Those significant at the 1 per cent level are double underlined.

FIGURE 1 ***MEAN PERSONALITY SCORES FOR LEADERS AND NON-LEADERS***

THE UNIT OF COMPARISON IS THE STANDARD DEVIATION OF THE DISTRIBUTION OF NON-LEADERS' SCORES. THUS, THE MEAN ELECTED LEADER SCORE ON FACTOR F IS ·83σ UNITS HIGHER THAN THE MEAN SCORE FOR GROUP MEMBERS WHO WERE NEVER ELECTED LEADER.

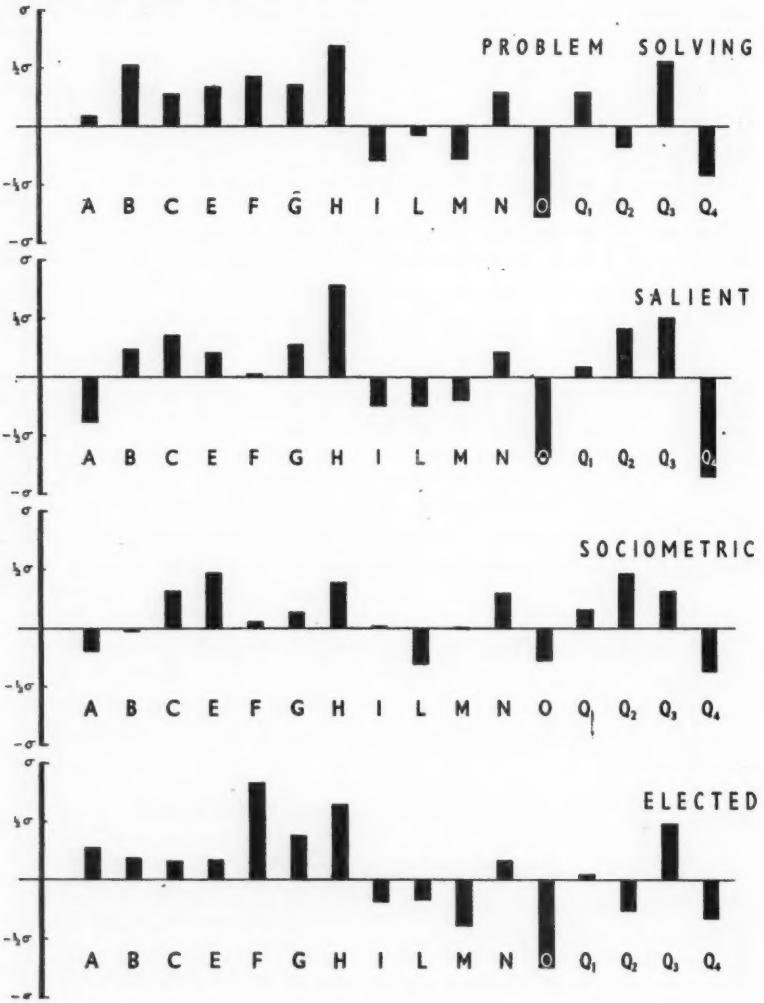


TABLE II *SIGNIFICANT PERSONALITY DIFFERENCES BETWEEN LEADERS AND NON-LEADERS FOR FOUR CRITERIA OF LEADERSHIP*

	1. Problem solving Leaders	*2. Salient Leaders*	*3. Sociometric Leaders*	*4. Elected Leaders*
A	—	—	—	—
B	4·2	—	—	—
C	—	2·2	—	—
E	—	—	3·1	—
F	2·5	—	—	4·6
G	3·1	—	—	3·6
H	3·1	3·4	—	3·2
I	—	—	—	—
L	—	—	—	—
M	—	—	—	−2·5
N	—	—	1·9	—
O	−2·2	−2·4	—	−2·7
Q_1	—	—	—	—
Q_2	—	—	2·4	—
Q_3	3·4	2·2	2·0	2·1
Q_4	—	−2·2	—	—

Note: 1. Differences are reported as Critical Ratios. Only differences which are significant at or beyond the 5% are shown.
2. The negative sign preceding a C.R. indicates that the leader score mean is lower than that of the non-leader mean.

On the other hand, special criteria of leadership selectively emphasize some factors (e.g., B, E, F, and O) much more than other criteria, and in the case of factor A and Q_2 there are even differences in what is the predisposing direction of endowment for leadership according to the concept of leadership used.

From what has been established about the nature of these personality factors in personality research generally, it is comparatively easy to see why these particular dimensions of personality should have such significant relations with leadership.

In regard to the three most consistently differing factors—H, O, Q_3—which are generally near, or well beyond, the 5 per cent level of certainty, the explanation is clear enough. The timid, withdrawn and hesitant behavior associated with H— would certainly militate against leadership. The anxious, worrying, cautiousness in dealing with people associated with O + would not inspire confidence from others. The absence of the will characters, the stability of purpose, and organizational precision, associated with Q_3 would not permit a person of otherwise suitable temperament to see his decisions through and to organize the group with consistency and a high degree of planning.

The fourth factor which is demanded in all forms of leadership and which reaches levels of significance almost equal to the above is G, or super-ego

strength. In so far as conscience may be said to be the "will of the group",—a regard for supernatural values—the selection of leaders with high G represents a dynamic gain for the group.

In discussing next those factors which differ appreciably for the different forms of leadership, it is necessary to review in some detail the differing conditions of selection. The reader may be reminded that the four categories of leaders (p. 496–7 above), are, 1. those observed to give most "leadership act" interaction with the group. These we might call "problem solvers" (No. 1 (252) above; p. 496); 2. Those noted by observers to be most influential on special occasions in affecting a group performance (No. 2 (253) above; p. 496). These we might call "salient leaders"; 3. those reported to be leaders by group members in retrospective checking. These we shall call "sociometric leaders" (No. 3 (112) above; p. 497); and 4. those actually elected by the groups as leaders in the election situations. These are simply "elected leaders" (No. 4 (413) above; p. 497). The significant differences between kinds of leaders is presented in *Table III.*

TABLE III *CRITICAL RATIOS FOUND BETWEEN THE MEAN PERSONALITY TRAIT SCORES FOR LEADERS SELECTED BY DIFFERENT CRITERIA*

	Method of Selection						
	1=Problem Solving	*2=Salient*		*3=Sociometric*		*4=Elected*	
Personality Trait	*Paired Differences*	1 : 2	1 : 3	1 : 4	2 : 3	2 : 4	3 : 4
A	—	(1·60)	—	—	(1·5)	2·1	(1·8)
B	—	2·3	3·2	2·7	—	—	—
C	—	—	—	—	—	—	—
E	—	—	—	—	(−1·5)	—	(1·7)
F	—	2·6	2·7	−2·6	—	−5·0	−5·0
G	—	—	—	—	—	—	—
H	—	—	—	—	—	—	—
I	—	—	—	—	—	—	—
L	—	—	—	—	—	—	—
M	—	—	—	—	—	—	—
N	—	—	—	—	—	—	—
O	—	—	—	—	—	—	(1·7)
Q_1	—	—	—	—	—	—	—
Q_2	—	−2·3	−2·7	—	—	2·8	3·2
Q_3	—	—	—	—	—	—	—
Q_4	—	—	—	—	—	—	—

Note: 1. All critical ratios which reach the 15 per cent level of significance (1·46) are shown in this table. It will be recalled that a C.R. of 1·96) is significant at the 5 per cent level and 2·58 at the 1 per cent level.

2. A minus sign preceding the C.R. indicates that the left hand (lower numbered) criteria has a lower mean (*Table I*).

For the selection of Problem Solvers (leader type 1), judged by the

"quantity of observed leadership", a high score is dependent upon the frequent appearance, at least momentarily, of influential behavior in a wide variety of situations, but not necessarily a large amount, or a continuing influence in any one situation. Thus, it makes good sense that here the *level of general ability* (factor B) shows more discrimination of problem solvers than of other leaders. Similarly, the spontaneous fluency of ideation and impulsiveness associated with Surgency (factor F) could lead to frequent bits of influential behavior, while the driving, persistent, group-concernedness associated with the super-ego-like Character Integration factor (G) would be expected to show a significant association through interest in the problems of the group.

Criterion 2, "Salient Leadership", on the other hand, was meant to select the people who, in at least one specific situation, evidenced reasonably clear and persistent leadership: these people were not merely influential, but the "most influential" for the whole situation. The predominant influence had moreover to be reasonably clear, in order for the two observers to agree upon it. Since leadership by this criterion must have involved stepping into and holding the center of attention for fairly long periods of time, it is not surprising that it is here that the separation is greatest for the adventurousness-shyness (H) factor and significant for both of the measures of anxiety (O and Q_4). It is perhaps surprising that the dominance (E) difference is not more marked, at least if we give any heed to one popular, over-simplified view which immediately associates dominance and leadership.

The sociometric criterion, consisting of weighted retrospective reports of leadership, is more similar in pattern to criterion 2 than to the others, indicating, we believe, a close correspondence between the salient and the sociometric definition of leadership (see *Table III*). It would appear, however, that the group member, in *recalling* his leaders was *more* influenced by dominance behavior (E), and less by adventurousness (H), and the "anxious, worrying, apprehensiveness" measured by factor O, than are observers on the lookout for behavior which influences what the group does. Furthermore, it should be noted that there is essentially *no* difference between the *level of ability* for leaders and non-leaders as selected by this criterion. If, as we have supposed, the intelligence measure can be taken as a rough index of effectiveness in problem solving and in the influencing of group behavior, then we are led to the interesting hypothesis that the sociometrically selected leader is not so much influential as he is attention-getting and dominating. He is the person who is recalled as having captured the group members attention, but he is not necessarily the person who did the things which were decisive with regard to the course of the group's progress.

Turning last to the elected leaders we notice that the chief way in which they differ from leaders by other criteria is on factor F, surgency. It is particularly remarkable that, while a high endowment on this trait is useful in securing election, it bears almost no relation to who is selected when the

group member is asked to recall "Whom do you judge to have been the leaders of this group throughout these meetings?" Indeed, if we could partial out the responses to this question which simply listed the elected leaders, it might be that, by the sociometric criteria, the low F person is in fact recalled as exercising more influence.

While the differences between leaders and non-leaders in factors A and Q_2 (Friendly Cyclothymia and Independent Self-sufficiency) are not significant, some of the differences *between leaders* selected by the different criteria are. Thus, it would appear that the warm friendly characteristics associated with A+, and the willingness to go with the group and obtain major dynamic gratification from interpersonal stimulation associated with $Q_2(-)$ make for election and for group influential behavior over a wide spectrum of situations. But, to an even greater extent the persistence in inner principles and lack of concern for the wishes of peers, associated with A− and Q_2+, (perhaps well illustrated by Woodrow Wilson) appears to be associated with Criteria 2 and 3.

These differences call for further examination in terms of (a) the relative satisfaction which group members received from the A−, Q_2+, in comparison with the A+, Q_2- leader, and (b) the dimensions of syntality that are altered by the two types of leaders. In the absence of further analysis of data, we would hypothesize that, because he will be more successful, the A−, Q_2+ leader will, in the long run, prove more popular in groups where the primary problem is to attain a difficult goal, but less successful and popular in groups and situations where the means of goal attainment are simple but the maintenance of contentment and morale of members is difficult. Similarly, the A−, Q_2+ leader will probably have his greatest effect upon the syntality dimensions related to group performance, while the A+, Q_2- leader will more likely influence the dynamic and interactional dimensions of structure and behavior.

While no direct measurement of the effect of leadership upon group productivity is as yet available, it seems probable to us that the observer ratings, in particular Criterion 2, will more closely reflect changes in productivity. Extrapolating from this assumption, and by reference to *Figure 1*, we may predict that the leader who had greatest influence upon the syntality of the group, will characteristically be much more adventurous (H+), less anxious (Q_4- and O−), more emotionally mature (C+) and more persistent and willful (Q_3+) than the typical member of the population from which he is drawn.

V. SUMMARY

1. Basically there are two ways of designating and measuring leaders: (a) by interaction observations and (b) by syntality change observations.

2. Although (b) is not yet fully practicable we have taken four opera-

tional definitions of leadership which range from (a) toward (b) and measured the personalities of individuals picked out by each, contrasting them with one another and with non-leaders.

3. The experimental situation was one in which 34 groups of 10 young men in each performed for 3 sessions of 3 hours each in a wide variety of common small group performances (e.g., committee decisions, construction, tug-of-war, cryptogram solving). The 13 syntality dimensions, factored out of 150 measures of external behavior and internal, structural, interactional observations, were not directly used here.

4. The four leadership categories may be designated:

(i) *Persistent Momentary Problem Solvers*, picked for frequency of brief acts of leaders. The equivalent of these in a role-structured group would probably be the "technical leaders". This is an interaction count. This comes near to defining the leader by his persistent influence on syntality.

(ii) *Salient Leaders* picked by outside observers as most powerfully influencing the group syntality on at least one total situation and occasion. This, like (i) above, comes near to designating leaders by group performance.

(iii) *Sociometric Leaders*, picked by frequency of subjectively reported perception of leadership by group members. This is a subjective, sociometric, rather than a behavioral, interactional, source of evidence on structure, and has nothing demonstrable to do with syntality change.

(iv) *Elected Leaders*, picked by voluntary election after experience. A structural measure interactionally based.

5. Comparing leaders with non-leaders, differences significant at the 1 per cent and 5 per cent level were found for factors G, O, H, and Q_3 (for leaders +, −, +, and +). These, and lesser differences on factors C, E, N, and Q_4, are in the same direction for all four classes of leaders.

6. In terms of the number of significant critical ratios, the greatest resemblance exists between salient and sociometric leaders and the least between sociometric and elected leaders.

7. The differences between leader types which have a significance at the 5 per cent level or beyond are that Problem Solvers are higher in General Intelligence, and Elected Leaders are higher in Surgency.

8. Four formulae, in the form of optimum factor patterns, may be gathered from the above tables, for selecting leaders according to these four conceptions. Any of these can be put alternatively into specification equation form, in order to utilize the 16 P.F. measures to give a simple prediction of leadership effectiveness. We have chosen to illustrate this by the "Elected Leaders" pattern, for this is probably of the widest practical application. From the above differences of means bi-serial correlations were worked out for each personality factor in relation to leadership. (As McNemar, for example, points out, the assumption that the dichotomized variable is continuous generally makes best sense, and it certainly does so in this case, where we are aiming to predict a continuous variable of "fitness for leadership"

which may be cut at any point according to the number of leaders required from the given population.)

The equation works out at:

$$P_L = \cdot16A + \cdot11B + \cdot09C + \cdot10E + \cdot50F + \cdot22G + \cdot35H - \cdot13I - 10L$$
$$- \cdot21M + \cdot09N - \cdot27O - \cdot16Q_2 + \cdot30Q_3 - \cdot28Q_4$$

Using this as a specification equation assumes that the factors are uncorrelated (this is true for the ideal personality factors but only approximately true for the present means of measuring them—The 16 P.F. Questionnaire), and that linear relations exist between the criterion and the personality factor endowments. This is a rather potent specification equation, perhaps because it covers so wide an array of personality aspects: for it accounts for 82 per cent of the variance of the criterion and gives a multiple correlation of ·91. It remains to be seen 1. how well this will stand up to cross validation with other groups; and 2. when more refined measures eliminate the overdetermination introduced by the neglect of slight correlations between the factors mentioned above. However, as it stands it appears the highest prediction of leadership that we have encountered in the literature.

BIBLIOGRAPHY

1. Carter, L., Haythorn, W., Shriver, B., and Lanzetta, J. "The Behavior of Leaders and Other Group Members", *J. Abn. Soc. Psychol.*, 1951, *46*, 589–595.
2. Cattell, R. B. "Concepts and Methods in the Measurement of Group Syntality", *Psychol. Rev.*, 1948, *55*, 48–63.
3. Cattell, R. B. "r_p and Other Coefficients of Pattern Similarity", *Psychometrika*, 1949, *14*, 279–298.
4. Cattell, R. B., Saunders, D. R., and Stice, G. F. *The 16 P.F. Questionnaire*. Champaign, Ill.; Institute of Personality and Ability Testing, 1951.
5. Cattell, R. B. *Personality: A Systematic, Theoretical and Factual Study*. New York: McGraw Hill, 1950.
6. Cattell, R. B. "New Concepts for Measuring Leadership in Terms of Group Syntality", *Hum. Rel.*, 1951, *4*, 161–184.
7. Cattell, R. B. "On the Theory of Group Learning", *J. Soc. Psychol.*, 1953, *37*, 27–52.
8. Cattell, R. B., Saunders, D. R., and Stice, G. F. "The Dimensions of Syntality in Small Groups: I. The Neonate Group", *Hum. Rel.*, 1953, *6*, 331–356.
9. Cattell, R. B., and Stice, G. F. *The Psychodynamics of Small Groups*, mimeographed report: Hum. Rel. Br., ONR, 1953.
10. Cattell, R. B. *The IPAT Humor Test of Personality*. Champaign, Ill.; Institute of Personality and Ability Testing, 1952.

11. CATTELL, R. B., and ANDERSON, JEAN. "The Measurement of Personality and Behavior Disorders by the I.P.A.T. Music Preference Test", *J. Appl. Psychol.*, 1953, *37*, 446–454.
12. GIBB, C. A. "The Principles and Traits of Leadership", *J. Abn. Soc. Psychol.*, 1947, *42*, 267–284.
13. GIBB, C. A. The Emergence of Leadership in Small Temporary Groups of Men. Ph.D. thesis, Univ. of Ill., 1949.
14. GIBB, C. A. "Leadership", in LINDZEY, G. (ed.) *Handbook of Social Psychology*, in press.
15. HEMPHILL, J. K. *Situational Factors in Leadership*. Ohio State Univ., Personnel Res. Bd., 1949.
16. JENNINGS, H. H. *Leadership and Isolation*, New York: Longmans Green, 1950.
17. STOGDILL, R. M. Leadership, Membership and Organization", *Psychol. Bull.* 1950, *47*, 1–14.
18. THURSTONE, L. L. *Multiple Factor Analysis*. Chicago: Chicago Univ. Press, 1947.

BIOGRAPHICAL NOTE

A biographical note on DR. RAYMOND B. CATTELL appeared in *Human Relations*, Vol. VI, No. 4, together with a note on DR. GLEN F. STICE. A further note on DR. STICE is included in this issue, in order to bring the notes up to date.

DR. GLEN F. STICE received his Ph.D. from the University of Illinois in 1951. He is now a Research Associate in the Department of Psychology, University of Illinois, where he is doing research in the area of personality assessment and social psychology. Prior to this he was with the Combat Crew Training Research Laboratory in Texas. His major research interests are problems of attitude definition and social group structure.

GROUP STANDARDS, COHESIVENESS, AND PRODUCTIVITY[1]

LEONARD BERKOWITZ

In a recent paper, Schachter, Ellertson, McBride, and Gregory (3) have demonstrated that the attractiveness of a group for its members (the group's cohesiveness) is not necessarily directly related to the level of group productivity. Other factors must intervene to provide a goal for the group, i.e., high or low production. Cohesiveness then exerts its influence by making for greater acceptance of this goal. In the words of the above-mentioned writers: "If the group attempts to influence its members to increase production, high cohesive groups should be more productive than low cohesive groups. If the group attempts to decrease production, low cohesive groups should be more productive than high."

The results of the experiment by Schachter and his colleagues support this prediction. The subjects in the cohesive group tended to accept more readily attempts to influence them to increase or decrease their production than did the subjects in the less cohesive groups. However, since the experiment was concerned only with the immediate success or failure of the attempts at influencing the production of the subjects, it did not determine whether the influence would persist when communication between the group members ceased, i.e., when pressure was no longer applied.

The present experiment is concerned largely with the question of the persistence of the influence. Through communication, the subjects are not only asked to work harder or to slow down in their rate of production, but they also learn the attitudes of their co-workers with regard to the designated group task. In other words, they learn what the group standard

1. I should like to express my appreciation to Dr. Robert L. Hall and Dr. William Haythorn for their very helpful suggestions toward the improvement of the research procedure; to Airmen Donald Bender, John Bowers, Jeffrey Gore, Philip Greco, William Hooper, and Rex Norgaard for their assistance in conducting the experiment; and to Mr. Lester Harrell of Trinity University, San Antonio, for his generous help in recruiting subjects and arranging for the use of Trinity University facilities. The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development program. The opinions and conclusions in this report are those of the author. They are not to be construed as necessarily reflecting the views or endorsement of the Department of the Air Force.

is. The concept "group standard" is a convenient abstraction by which we mean the expressed and shared attitudes of the group members toward their task.[2] A standard of high production, then, refers to attitudinal consensus favorable to the designated group task, while a standard of low production indicates that most of the group members share an attitude less favorable to the achievement of the designated group task. Our hypothesis is that, even after direct pressure is no longer placed upon them, members of high cohesive groups will tend to behave in closer conformity to the perceived group standard than will members of less cohesive groups. Where the standard is for high production, the high cohesive group will tend to increase their production; where it is for low production, the high cohesive groups will tend to decrease production, or at least will not increase it as rapidly as the low cohesive groups. This implies not merely a temporary acceptance of influence but a more permanent conformity to a perceived attitude.

The classic Sherif experiment (4) has demonstrated that the group's influence upon the individual's perceptual judgments can persist up to the time when the individual is alone, i.e., when there is no communication among the group members. In the present experiment we are interested in determining whether the group's influence upon the individual's productivity can persist when there is no communication among the group members, and whether this varies with the cohesiveness of the group.

A secondary purpose for this investigation is to test some of the incidental findings in the above-mentioned experiment. Although, in the Schachter experiment, there was apparently a significant interaction between the direction of influence—i.e., toward raising or lowering production—and cohesiveness, in determining productivity, the high production and low production conditions were not affected equally. The subjects who were asked to decrease their production, and who were in cohesive groups, appeared to be influenced more successfully than the subjects who were asked to increase their production, and who were in cohesive groups. This result is worth checking because of its unexpected nature, and because of the attention given it by the above writers.

PROCEDURE

The experiment conducted by Schachter and his colleagues, provided the model for the present experiment. As in their design, there are four conditions[3]:

1. Standard of high production, high cohesiveness.

2. Attitudinal consensus is an important part of the concept of group standard. There is some evidence that the power of the group to define the socially proper behavior for the individual member is a function of the perceived agreement among the group members (1).

3. Schachter, *et al.*, termed their two major conditions "positive or speed-up induction" and "negative or slow-down induction".

2. Standard of high production, low cohesiveness.
3. Standard of low production, high cohesiveness.
4. Standard of low production, low cohesiveness.

With the exception of the variations made in the manipulation of the group's standards and cohesiveness, the procedure was identical in all experimental conditions.

The subjects were males recruited from ROTC, government and economics classes. There were 9 subjects in each of the two high production standard conditions and 8 subjects in each of the two low production standard conditions, a total of 34 subjects.

Two to four subjects were scheduled in any one hour. As each subject arrived, he was placed in a room by himself and asked to wait "until his group could be started". The experimenter then went to each of the rooms and described the group's task in a brief fashion. Information making for higher or lower cohesiveness was also provided at this time, as is described below. This procedure was followed for each of the subjects. The experimenter then made the rounds again, introducing two of his assistants as the subject's "co-workers". (The small size of the school made it necessary to use outside "stooges" since the subjects frequently knew each other.) The task was now demonstrated in more detail.

The subjects were told that their group was going to make ash-trays "on an assembly-line basis". One of the three was to be the "cutter" whose job was to cut discs $2\frac{7}{8}''$ in diameter from large desk blotters. Circles were already drawn on the blotters. The second man was to take the role of "painter" who would work independently of the "cutter" and stencil designs onto discs of paper the same size as the discs of blotter. It was explained that this paper had been pre-cut, and the "painter" would be given a sufficient supply. The third man, the "assembler", would receive the finished products of the first two, and would assemble the completed ash-trays by pasting one design onto each disc of blotter and then pasting these onto the back of a suitable glass receptacle. In explaining the task, the experimenter emphasized that quantity of production was most important, although quality was also important.

When all three tasks had been demonstrated, the experimenter implied that the experiment was concerned with studying communication in situations where the group members could not communicate "naturally" (in a face-to-face manner) but had to communicate "artificially"—as "in the latest jet-bomber". The subjects were told that artificiality of communication would be ensured by having each working in a separate room, but that they could write notes to each other. It was explained that a messenger—the experimenter or one of his assistants—would come by at regular intervals to collect their finished productions, and to collect and deliver notes to and from members of the team. This procedure, they were told, would last twenty or twenty-five minutes, in which time they could write as many

notes as they wished. After this time the notes would no longer be delivered and they would work for another period without communicating; the total experiment would be over in "about three-quarters of an hour". In conclusion, the experimenter said he would return the subjects to their respective rooms and then assign the three jobs to them "at random, in order to eliminate any possible bias in the distribution of jobs".

After this introduction, the experimenter and his two "stooges" left to repeat this procedure with a second subject. A few minutes after their departure a messenger arrived, explained that the subject was going to be the cutter, supplied him with desk blotters, scissors, pencil, and a pad of note-paper, and set him to work. He returned every four minutes to pick up the subject's finished production, the discs of blotters, and any notes that he had written. Prewritten notes were delivered to him. These notes furnished the means by which the standard of production was manipulated. The subject cut discs from the blotters for forty minutes. Then, after filling out a 4-item questionnaire designed to assess the attractiveness of the group for him, he was brought to the experimenter. The purpose of the experiment was then discussed and the various deceptions were explained.

The Manipulation of the Variables

Cohesiveness. High and low cohesiveness were produced by manipulating the liking of the subjects for their "co-workers". This manipulation began during the recruiting when the experimenter met with several classes and called for volunteers. In his speech to the classes, he said it was important to have groups of people "who would be sure to like each other". He told them that "on the basis of recent findings it was now possible to tell what kinds of people would probably like each other and what kinds of people would probably not get along well with each other". For this "matching" the volunteers filled out a questionnaire ostensibly designed to get personality information. There were eight multiple-choice items in this questionnaire.

When the subjects arrived for the experiment and were privately interviewed, they were randomly assigned to either the high or low cohesive conditions. In the high cohesive condition the subject was told he was in an extremely congenial group and that "there was every reason to believe" that he would like the people working with him and that they would like him. The low cohesive subjects were told that "due to scheduling difficulties it had been impossible to bring together a congenial group", and that "there is no reason to think that you will like the people working with you or that they will like you. As a matter of fact, you may get to dislike them over time—there isn't enough in common in your personalities."[4]

Two of the experimenter's assistants served as "stooges" in the great

4. This is a somewhat stronger statement than is common in the low cohesive manipulation. Frequently experimenters using this technique have concluded with the statement that "there is no reason to think that you will like . . ."

majority of cases. In those instances where an assistant could not participate in the experiment at a certain hour, he was replaced by another volunteer. In general, each assistant served as a "stooge" in as many low cohesive as high cohesive groups.

At the end of the experiment, each subject filled out a cohesiveness questionnaire designed to test the success of the manipulation. There were four Likert-type items in this questionnaire:

1. How did you like your team?

2. Do you believe that your team-mates had attitudes concerning the group's job that were similar to your own attitudes?

3. If you were taking part in another experiment, how much would you like to work with these same fellows?

4. If it involved any effort on your part, would you care to know the fellows better?

The items seemed to discriminate equally well between the high and low cohesiveness conditions, and were combined to yield a single cohesiveness score.

The Group Standard. Cues as to the attitudes of the subjects' "co-workers" were provided by the prewritten notes. There can be little doubt, judging from their responses to these notes as well as from their statements upon the completion of the experiment, that the very great majority of the subjects accepted the notes as coming from their "co-workers". Each subject received seven notes. The first two notes—delivered 4 and 8 minutes after

Time interval from commencement of experiment	*Notes issued during experiment*	
	To all subjects	
4 minutes	"I wonder who will use these ash-trays" (assembler)	
8 minutes	"I wonder who thought up this design" (painter)	
	To high production subjects	*To low production subjects*
12 minutes	"Let's try to set a record" (painter)	"You're getting way ahead of me—relax" (assembler)
16 minutes	"The assembler says he is twiddling his thumbs. Let's give him plenty of work" (painter)	"We don't have to wear ourselves out making these ash-trays" (painter)
,,	"I'm running a little short of material" (assembler)	"Take it easy, I'm tired" (assembler)
20 minutes	"Let's keep up a fast, steady clip" (assembler)	"We've done a lot of these things. Let's take it easy now" (assembler)
,,	"This is the last note, so don't slow down" (painter)	"Painting these things takes more time than you think—let's slow down" (painter)

the subjects began work—were neutral in content and did not attempt to influence production. This provided a baseline against which the subject's later production could be compared. In the next 12 minutes, the notes attempted to increase or decrease the subject's rate of production by providing evidence as to the attitudes of the "co-workers" concerning the designated group task. Here, too, the post-experiment comments of the subjects indicate the success of the manipulation. Many of the subjects reported that they believed their co-workers to be "lazy" or "very eager". After 20 minutes, notes were no longer delivered and the subjects worked until 40 minutes had passed from the time they had begun cutting. The specific notes and the times at which they were delivered were as given on page 513.

RESULTS

Table I presents the mean cohesiveness score in each of the four conditions. Analyses of variance reveal that in both the low and high production conditions the high cohesiveness subjects were significantly more attracted to their "co-workers" than the corresponding low cohesiveness subjects. The cohesiveness manipulation can thus be regarded as being successful.

TABLE I *MEAN COHESIVENESS SCORE IN EACH OF THE FOUR EXPERIMENTAL CONDITIONS*

	High Production		*Low Production*	
	High Coh.	*Low Coh.*	*High Coh.*	*Low Coh.*
N	9	9	8	8
Mean	17·4	14·6	16·0	11·2
F-Ratio	6·23		7·82	
P	·05–·01		·05–·01	

The mean number of units of production in four time intervals in each of the conditions is shown in *Table II* and *Figure 1*. It can be seen that there are differences in the initial rate of production—the rate per 4-minute interval is lower in the Low Production, Low Cohesive condition than that in any other condition. Because of these conditions and individual differences, the most meaningful measure for our purposes is not the number of units *per se*, but rather the average difference (the increase or decrease) between the subject's initial production and the production after cues were provided as to group standard.

Table III shows the analysis of variance in the four conditions using the mean number of units difference between the subjects' production per 4 minutes in the baseline period (0–12 minutes) and their production per 4 minutes in the standard setting period (12–24 minutes). (A constant of

TABLE II *THE MEAN NUMBER OF UNITS OF PRODUCTION PER FOUR-MINUTE INTERVAL AND MEAN CHANGE IN PRODUCTIVITY FROM THE BASELINE PERIOD IN EACH OF THE FOUR EXPERIMENTAL CONDITIONS*

Time Period	*High Production Standard*		*Low Production Standard*	
	High Coh.	*Low Coh.*	*High Coh.*	*Low Coh.*
0–12 min. (baseline)				
Mean Production	6·00	6·38	6·45	4·94
12–24 min. (standard setting)				
Mean Production	8·21	6·87	5·36	4·88
Mean Change	2·21	·49	−1·09	— ·06
P of Difference	·05–·01		·05–·01	
24–32 min.				
Mean Production	9·39	8·00	7·00	7·06
Mean Change	3·39	1·62	·55	2·12
P of Difference	·05–·01		·10–·05	
32–40 min.				
Mean Production	10·67	8·44	7·06	7·25
Mean Change	4·67	2·06	·61	2·31
P of Difference	·05–·01		·05–·01	

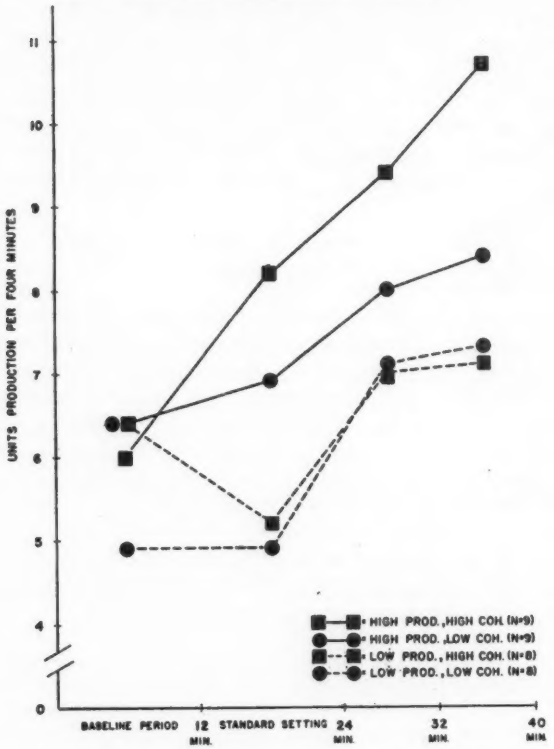
5 was added to the score of each subject in order to eliminate negative values—6 of the 8 subjects in the Low Production, High Cohesive condition immediately decreased their rate of production.)

TABLE III *ANALYSIS OF VARIANCE OF CHANGES IN PRODUCTIVITY FROM BASELINE PERIOD TO STANDARD SETTING PERIOD IN THE FOUR EXPERIMENTAL CONDITIONS*

Source of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*	*F*	*P*
Total	33	88·84			
Between	3	48·02	16·01	11·77	·001
Standard	1	29·56	29·56	21·74	·001
Cohesiveness	1	1·13	1·13	—	—
Interaction	1	18·33	18·33	13·5	·001
Within	30	40·82	1·36		

These results tend to confirm the results obtained by Schachter and his colleagues (3). There is a significant interaction between the group standard and group cohesiveness in determining productivity in the time period during which the group standard is established.

FIG. 1
THE MEAN NUMBER OF UNITS OF PRODUCTION PER FOUR-MINUTE INTERVAL DURING SEVERAL PERIODS UNDER EACH OF THE FOUR EXPERIMENTAL CONDITIONS



The next hypothesis was that these differences would persist after the standard-setting period. The subjects are now aware of their "co-workers'" attitudes, and, if the group is cohesive, should still try to behave in conformity with the group standard. *Table IV* presents the trend analysis of variance for the differences between production in the baseline period and production in each of the subsequent three periods. In general, it can be seen that the significant interaction between cohesiveness and group standards persists over all the time periods. The error term used here is the mean square of the production within conditions. Furthermore, time does not affect this relationship nor the relationships between cohesiveness and productivity and standards and productivity. The interaction between time and the various experimental conditions are all insignificant using the interaction of time and the pooled subjects as the error term.

TABLE IV *TREND ANALYSIS FOR THE CHANGES IN PRODUCTIVITY FROM THE BASELINE PERIOD TO THE STANDARD SETTING PERIOD AND TO THE TWO INTERVALS, 24–32 MINUTES, AND 32–40 MINUTES, AFTER THE CESSATION OF COMMUNICATION*

Source of Variation	*Sum of Squares*	*DF*	*Mean Square*	*F*	*P*
Between Conditions	151·01	3	50·34	10·73	·001
Standards	72·76	1	72·76	15·51	·001
Cohesiveness	4·36	1	4·36	0·93	—
Interaction: stand. × coh.	73·89	1	73·89	15·75	·001
Between Subjects in Same Condition	140·78	30	4·69		
Total between subjects	291·79	33			
Between Time Periods	70·81	2	35·40	32·78	·001
Interaction: time × conditions	5·25	6	0·88	0·81	—
Interaction: coh. × time	0·65	2	0·32	0·30	—
Interaction: stand. × time	2·68	2	1·34	1·24	—
Interaction: coh. × stand. × time	1·92	2	0·96	0·89	—
Interaction: pooled subjects × time	64·69	60	1·08		
Total Within Subjects	140·75	68			
Total	432·54	101			

We do not know the time limits within which the standards-cohesiveness interaction will be related to productivity. Over a longer period of time it is conceivable that the cohesiveness and standard setting manipulations will lose their effectiveness so that the production within all four conditions will tend to approach the same rate. However, there is no indication that this is happening within the duration of the present experiment. The differences between conditions are increased within the three time periods, but the relationships between productivity and the experimental variables are apparently not affected by the passage of the 28 minutes. The changes over time for the four experimental conditions are presented in *Figure 1*. (The units used in this figure are absolute units rather than deviations from the baseline period.)

DISCUSSION

Several experiments have demonstrated that the power of the group over the individual member—the extent to which it can influence his behavior—

is a function of the attractiveness of the group for that individual (2). In the present case the influencing factor was the group standard. Through communication, the subjects were not only asked to work harder or to slow down in their rate of production, but they also learned the attitudes of their "co-workers" concerning the designated group task. The high cohesive subjects increased or decreased their rate of production, depending upon the group standard, in immediate response to the communications addressed to them, and then continued, relatively speaking, to conform to the perceived attitudes of their fellow group members after intra-group communication had ceased. Thus, the influence of the group standard persisted in the high cohesive groups within the length of time investigated, and influenced the group members' production, even though pressure was no longer placed upon them to conform to the standard. It is reasonable to suppose that this influence will persist as long as the group remains cohesive and nothing happens to change the perceptions of the attitudes of the co-workers.

Schachter, Ellertson, McBride, and Gregory (3) did not obtain any significant differences between the high and low cohesive groups in the two high production conditions, and formulated some elaborate hypotheses concerning why cohesiveness should only make for the acceptance of pressure to decrease production. The present experiment, however, reveals statistically significant differences in the change from the baseline period between the two high production as well as between the two low production conditions in all time periods. This can be seen in *Table II*. It may well be, then, that the previously mentioned results are an artifact of the experimental task. The subjects in the High Production, High Cohesive condition might well have been motivated to increase their production by the perceived group standard. However, they may not have been able to raise their rate of production significantly above that in the corresponding low cohesive condition because they were already at the task ceiling. They could not do more because of the nature of the task—cutting large rectangles from heavy cardboard—even though their motivation may have been higher than that of the corresponding low cohesive subjects.

SUMMARY

Thirty-four male college students were randomly assigned to four conditions: two conditions having a group standard of high production but differing in group cohesiveness, and two conditions having a group standard of low production but also differing in group cohesiveness. Cohesiveness was manipulated by telling the subjects they would probably like their co-workers with whom they were supposedly matched, or that they might not like each other. In every case the "co-workers" were the experimenter's assistants. Though believing they were working on a co-operative group task—producing ash-trays—the subjects actually worked alone and all cut

out discs from desk blotters. The group standard was manipulated by delivering prewritten notes to them at fixed intervals giving them the impression that their "co-workers" were highly motivated toward the group task and wanted them to work hard, or were unfavorably disposed toward the group task and wanted them to slow down.

For the three experimental time periods—the standard setting period (12 to 24 minutes after the beginning of the experiment), the first persistence period (24 to 32 minutes), and the second persistence period (32 to 40 minutes)—there was found to be a statistically significant interaction between the group standard and the group cohesiveness in determining the change in productivity from the baseline period. Neither this interaction, nor the effect of either variable alone, changed during the experimental period as indicated by the results of a trend analysis.

These results are taken as corroborating the major hypothesis. Members of cohesive groups will tend to conform to the perceived group standard—by raising their production if the standard is for high production or lowering it, or not increasing as rapidly, if the standard is for low production—even after overt pressure to conform to the group standard is discontinued.

BIBLIOGRAPHY

1. ASCH, S., "Effects of Group Pressure Upon the Modification and Distortion of Judgments", in GUETZKOW, H. (Ed.) *Groups, Leadership and Men*, pp. 177–190. Carnegie Press, 1951.
2. FESTINGER, L. "Informal Social Communication", *Psychol. Rev.*, 1950, *57*, 271–282.
3. SCHACHTER, S., *et al.* "An Experimental Study of Cohesiveness and Productivity", *Hum. Rel.*, 1951, IV, 229–238.
4. SHERIF, M. "A Study of Some Social Factors in Perception", *Arch. Psychol.*, 1935, No. 187.

BIOGRAPHICAL NOTE

DR. BERKOWITZ received a B.A. from New York University in 1948, and a Ph.D. from the University of Michigan in 1951. As a research psychologist for the Crew Training Research Laboratory, Human Resources Research Center, Randolph Air Force Base, Texas, he is studying the social characteristics of bomber crews that are related to their effectiveness.

BOOK REVIEW SECTION

Owing to pressure of space the Book Review Section has been held over until the next issue.

STUDIES OF SOCIAL BEHAVIOUR

carried out in Australia by Staff and Students
of the Department of Psychology, University of Melbourne

SOCIAL STRUCTURE AND PERSONALITY

VOLUME ONE: IN A CITY

Edited by O. A. Oeser and S. B. Hammond

VOLUME TWO: IN A RURAL COMMUNITY

by O. A. Oeser and F. E. Emery

These two volumes are the report of a study made under the auspices of the Unesco tensions project in social structure and personal relations in Australia. The first deals with the city and the second with the rural community. They provide an important study of society and the techniques they present are already being applied in England and the United States.

The first volume concerns the phases through which a person passes in the course of his life, and the roles he assumes in response to varying social demands: as infant, school-child, adolescent, worker, and husband or wife. An overall picture emerges of Australian society and its way of life, and the attitudes of the citizen to such important themes as immigration and assimilation, the relation of social position and the dynamics of family life.

The second volume describes the background of the wheat-producing community, the perceptions of the social structure and the development of ideologies. A detailed study is made of the integrative social role of the school in relating the community to the wider society.

Demy 8vo. Volume One, 30s. net. Volume Two, 25s. net.

ROUTLEDGE & KEGAN PAUL

PSYCHIATRY

JOURNAL FOR THE STUDY OF INTERPERSONAL PROCESSES

CONTENTS OF NO. 2, VOL. XVII, MAY 1954

MABEL BLAKE COHEN, GRACE BAKER, ROBERT A. COHEN, FRIEDA FROMM-REICHMANN, AND EDITH V. WEIGERT: *An Intensive Study of Twelve Cases of Manic-Depressive Psychosis.*

JULES HENRY: *The Formal Social Structure of a Psychiatric Hospital.*

OSCAR LEGAULT: *Denial as a Complex Process in Post Lobotomy.*

ROBERT N. WILSON: *Poetic Creativity.*

SHIRLEY W. KAHN AND A. RODNEY PRESTWOOD: *Group Therapy of Parents as an Adjunct to the Treatment of Schizophrenic Patients.*

ELINOR W. DEMAREST AND ARTHUR TEICHER: *Transference in Group Therapy.*

Editorial Notes.

Brief Communications.

Book Reviews.

Published quarterly by the William Alanson White Psychiatric Foundation, 1711 Rhode Island Ave., N.W. Washington 6, D.C., U.S.A.

OCCUPATIONAL PSYCHOLOGY

Editor: ALEC RODGER

JULY, 1954 VOLUME 28, NO. 3

CONTENTS

Attitudes to Work — *by* L. S. HEARNSHAW

Satisfaction and Satisfactoriness: Complementary Aspects of Occupational Adjustment — *by* ALASTAIR HERON

Students' Progress in a Medical School — *by* A. E. MAXWELL

Occupational Psychology: Index, 1938–1953 — *Compiled by* P. F. C. CASTLE

Book Reviews

Other Books Received

Annual Subscription 30 shillings

National Institute of Industrial Psychology, 14 Welbeck Street, London, W.1.

1954 HUMAN RELATIONS Vol. VII, No. 4, Pages 401—520

CONTENTS

	Page
STANLEY SCHACHTER, JOSEF NUTTIN, CECILY DE MONCHAUX, PAUL H. MAUCORPS, DIEDRICH OSMER, HUBERTUS DUIJKER, RAGNAR ROMMETVEIT, and JOACHIM ISRAEL. Cross-Cultural Experiments on Threat and Rejection ..	403
F. KRÄUPL TAYLOR. The Three-Dimensional Basis of Emotional Interactions in Small Groups	441
ORVILLE G. BRIM, Jr. The Acceptance of New Behavior in Child-Rearing	473
RAYMOND B. CATTELL and GLEN F. STICE. Four Formulae for Selecting Leaders on the Basis of Personality	493
LEONARD BERKOWITZ. Group Standards, Cohesiveness, and Productivity	509

November, 1954 World List of Abbreviations: HUM. REL.

Printed in Great Britain by Butler & Tanner Ltd., Frome and London